The Philippine Statistician, 2007
Vol. 56, Nos. 1-2, pp. 1-8

The Design and Analysis of Pivotal Clinical Trials To Assess the Efficacy of Drugs to Treat Panic Disorder

Karl E. Peace[1]

ABSTRACT

US Federal Regulations require at least two pivotal proofs of efficacy clinical trials (typically Phase III) to support approval of a drug for a specific indication. In the clinical development of drugs to treat panic disorder, one of these trials may incorporate a forced titration design while the other may incorporate a flexible titration - according to response design. Standard analysis approaches include treatment group comparisons at each follow-up visit in terms of variables reflecting domains of the panic condition. Such designs pose challenges as to: (1) characterizing the dose response profile across both studies; {2) characterizing duration of effect; (3) characterizing withdrawal effects; and (4) the choice of appropriate statistical analysis methods. Standard approaches to meeting these challenges, as well as reasonable alternatives are presented.

Key Words: Pivotal, Clinical Trials, Panic Disorder, Duration of Effect, Withdrawals

I. INTRODUCTION

Patients with panic disorder experience sudden, unexpected episodes of intense apprehension or terror (a panic attack) with no apparent stimulus. Typical signs and symptoms include hyperventilation, tachycardia or palpitations, chest pain, sweating, trembling, and sensations of smothering or choking. Patients with the disorder may also experience blurred vision, weakness, or feelings of unbearable dread or terror. Many become so demoralized they are unable to leave their homes [1]. Panic disorder may occur with or without agoraphobia. Agoraphobia is fear of places where help may not be available – such as crowded places or remote and isolated places.

Panic disorder with or without agoraphobia is common. The overall 12-month and lifetime prevalence rates are 2.1% and 5.1% [2]. It is a severe condition in those affected; often requiring visits to emergency rooms.

US Federal Drug Regulations require at least two pivotal proofs of efficacy clinical trials to support approval of a drug for a specific indication [3]. For drugs not yet approved, pivotal proofs of efficacy trials occur in Phase III of the clinical development program. However they may occur post marketing as Phase IV trials when a company develops evidence to support labeling for additional indications for an approved drug.

[1] Fellow of the American Statistical Association, Georgia Cancer Coalition Distinguished Cancer Scholar, Senior Research Scientist & Professor of Biostatistics, Founding Director, Karl E. Peace Center for Biostatistics, Jiann-Ping Hsu College of Public Health, Georgia Southern University, PO Box 8015, Statesboro, GA USA 30460. Email address: kepeace@georgiasouthern.edu, peacekarl@cs.com

In the clinical development of drugs to treat panic disorder, one pivotal proof of efficacy trial may incorporate a randomized, forced titration dose response design while the other may incorporate a flexible titration (according to response) design. Standard analysis approaches include treatment group comparisons at each follow-up visit in terms of variables reflecting domains of the panic condition. Such designs pose challenges as to characterizing:

1. The dose response profile across both studies;
2. Duration of effect;
3. Characterizing withdrawal effects; and
4. The choice of appropriate statistical analysis methods.

Standard approaches to meeting these challenges, as well as reasonable alternatives are presented.

II. DESIGN OF TRADITIONAL PIVOTAL PROOF OF EFFICACY TRIALS

The **experimental, statistical design** for both the forced titration dose response trial (**FTDRT**) and the flexible titration according to response trial (**FTART**) is a completely randomized block design, with investigational centers as blocks. In addition, both are parallel, double blind and placebo controlled. Both trials [4, 5, 6] consist of 5 consecutive phases: a one week placebo baseline run-in period — after which patients diagnosed with panic disorder [7] and who satisfy protocol eligibility criteria are randomized; then an upward titration dosing period of up to three weeks; then a fixed dosing period of three to six weeks; then a downward titration dosing period of variable length -- depending on the patient's fixed dose; and finally a one week placebo washout period.

Both trials have approximately the same visit schedule: the initial screening visit; the baseline visit; weekly visits during the upward titration period; weekly or bi-weekly (depending on length) visits during the fixed dosing period; one visit during 'middle' of downward titration period and 1 at the end; and a final visit at end of the placebo washout period.

For the **FTDRT** (Figure 1), there may be up to six randomized dose groups: D_0, D_1, D_2, D_3, D_4, and D_5, where D_0 represents the placebo group, and D_i, i = 1, ..., 5, represent fixed doses of the drug under study. After randomization, 1/6 of the patients are on D_0 for the duration of the trial, and 5/6 are at D_1 for a fixed period of time, say T; then 4/5 of these move to D_2 for the period T and the other 1/5 stay at D_1 until the end of the fixed dose period; then 3/4 of those on D_2 move to D_3 for the period T and 1/4 stay at D_2 until the end of the fixed dose period; then 2/3 of those on D_3 move to D_4 for the period T and 1/3 stay at D_3 until the end of the fixed dose period; then 1/2 of those on D_4 move to D_5 for the period T and 1/2 stay at D_4 until the end of the fixed dose period; so that after a period of 5T (the upward titration period), all patients are at their randomized fixed dose, where they stay until the end of the fixed dosing period. It is noted that the randomization 'forces' upward titration of dose regardless of response, and packaging ensures blindness as to the identity of the dose.

For the FTART (Figure 2), there are two randomized groups (D_{v0}, D_{v1}), where D_{v0} represents the placebo group at variable 'strengths', and D_{v1} represents the drug under study

group (at variable strengths). The strengths available are usually the same as those used in the FTDRT, D_1, D_2, D_3, D_4, and D_5. After randomization, 1/2 of the patients are on placebo throughout the trial period, and 1/2 are on drug through the end of the fixed dosing period. All patients (in the placebo and drug groups) are at strength D_1 for the period T; those who respond (clinician's judgment) stay at D_1, and those who don't are moved to strength D_2 for an additional period of T; those who respond to D_2 stay, and those who don't are moved to strength D_3 for an additional period of T; those who respond to D_3 stay, and those who don't are moved to strength D_4 for an additional period of T; those who respond to D_4 stay, and those who don't are moved to strength D_5 for an additional period of T. After a period of 5T, all patients are at their 'optimal' fixed dose, where they stay until the end of the fixed dosing period. If a patient has not responded by the end of the fixed dosing period, the investigator has the option to treat the patient outside the protocol. It is again noted that packaging ensures blindness.

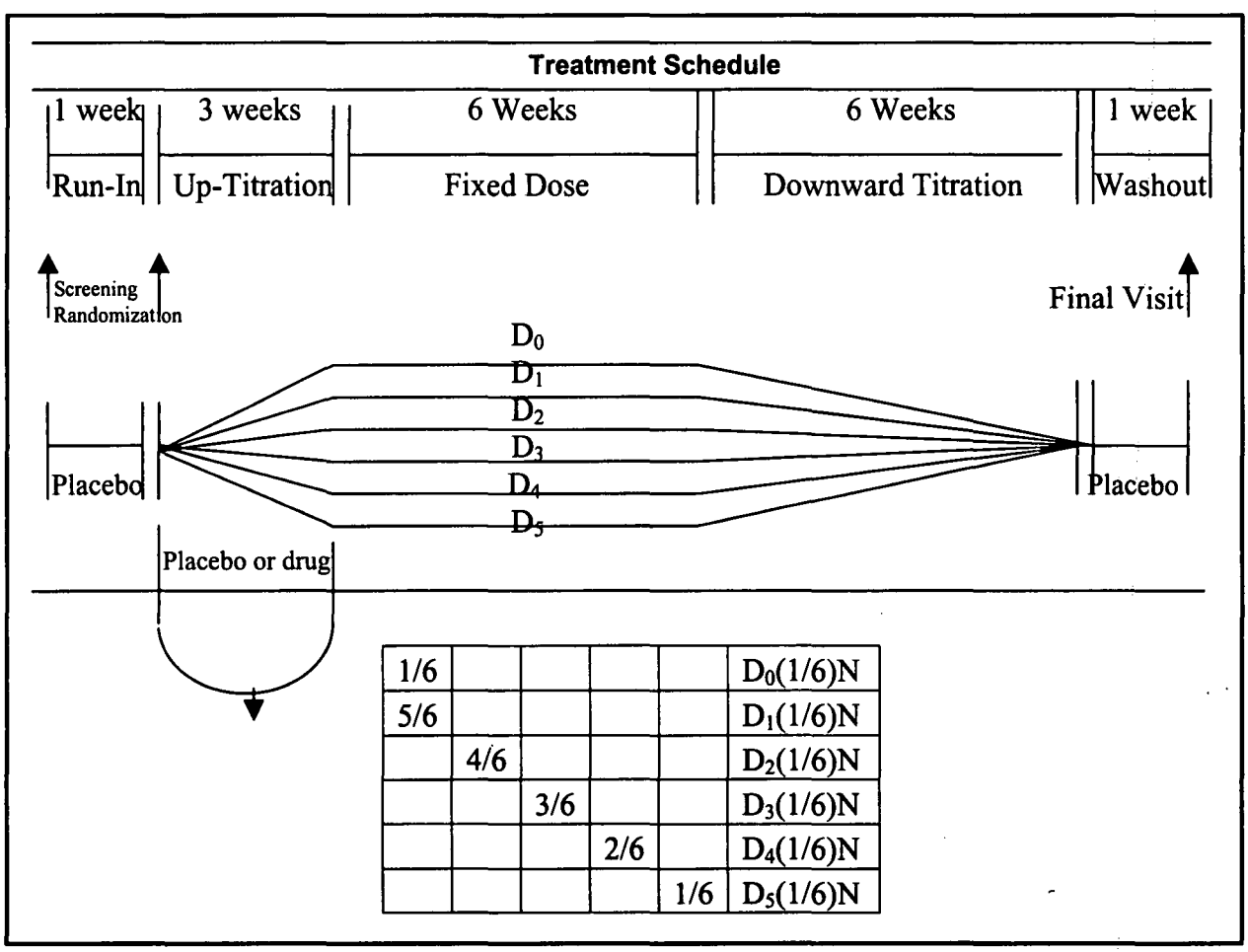

Figure 1: Forced Titration, Dose-Response Trial Schema

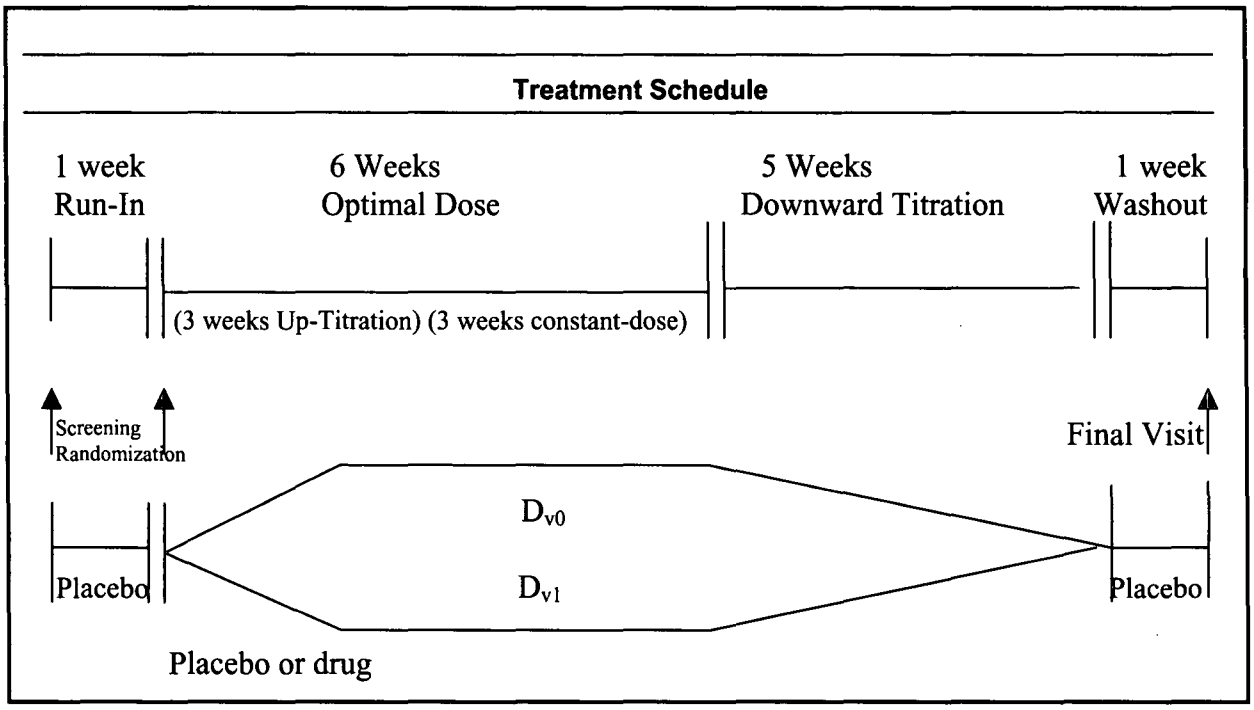

Figure 2: Titration According to Response Trial Schema

Efficacy variables or endpoints collected in such trials are many and reflect domains of the panic condition. **Primary efficacy measures** are: the number of panic attacks (**NPA**) between clinic visits and the clinician's global impression of severity (per a numeric rating scale) of the panic condition (**CGIS**), both recorded at each visit to the end of the fixed dosing period.

Secondary efficacy measures include: the clinician's global impression of change (**CGIC**) in the panic condition (recorded at post baseline office visits) and the patient's global impression of severity (**PGIS**) of his/ her panic condition. There are many other secondary measures, including: an assessment of activities of daily living, assessment of anxiety, items from the Hamilton Depression Scale, and data recorded on the patient's daily diary.

Obviously, **the objective** of each trial is to demonstrate efficacy of the drug. Since each trial is placebo controlled, a one sided alternative is appropriate [8] with a Type I error or false positive rate of 5%.

Although, many choose a power of 80% it is good practice for pivotal proof of efficacy trials to be designed with 95% power [9]. The **number of patients** required to participate in each trial would then be determined using these error rates to detect a pre-specified improvement (above placebo) in terms of the primary efficacy measures NPA and CGIS. For the FTDRT trial, one has to decide how dose response should be characterized. One choice would be to detect a non-zero slope of the linear component of the dose response curve. Another choice would be to detect a difference of δ (not greater than 20%) between a dose group and the placebo group. For the FTART trial, efficacy would be characterized as a difference of δ between the drug group and the placebo group.

III. TRADITIONAL STATISTICAL ANALYSIS METHODS

Analysis of efficacy is restricted to efficacy data collected to the end of the fixed dosing period. Data collected past this point may be summarized and analyzed in an attempt to characterize withdrawal effects. Traditional statistical analysis methods of efficacy measures are analysis of variance (ANOVA) or analysis of covariance (ANCOVA) methods. The choice of any particular analysis method should depend on whether the assumptions undergirding the validity of the method are appropriate for the behavior of the data [10]. Since the distribution of the NPA is usually skewed, ANOVA or ANCOVA of ranks or non-parametric methods such as Cochran-Mantel-Haenszel [11] should be performed. Since change from baseline within an intervention group is an index of the effect of the intervention received, change from baseline (CNPA) in the NPA and change from baseline (CCGIS) in the CGIS may be preferred as primary measures of efficacy instead of NPA and CGIS, respectively.

There is a randomization basis for inferences as to dose response from the FTDRT trial. A statistically significant slope of the dose response relationship over the six dose groups provides unequivocal evidence of a drug effect. This should be followed by step-down procedures to identify the minimum effective dose. This information will be helpful in the labeling of the drug once approved for marketing.

There is no randomization basis for inferences as to dose response from the FTART trial. The only valid inference randomization based inference is the pair wise comparison of drug group to placebo group.

IV. DISCUSSION

Since the FTDRT and FTART trial are multi-center, longitudinal with multiple endpoints the usual issues inherent in such trials have to be adequately dealt with by the analyst. These include: **interactions** – treatment group-by-center, treatment group-by-baseline factors such as disease severity or demographics, treatment group-by-time, how to handle **missing data**, and how to deal with the **impact of multiplicities** on overall conclusions.

In addition, for the **FTDRT** trial, it is unlikely that there will be a clinic visit at each time (T, 2T, ..., 5T) of forced upward titration, and consequently there will be no **efficacy data at some times of dose escalation**. Were patients to return to the clinic at each time of such dose escalation and data recorded just prior to the escalation, it would be possible (provided T is sufficiently long) to assess the incremental benefit of the next dose level (for a period of T) as compared to remaining at the current dose level. [The de-escalation or withdrawal period suffers this same criticism.] Therefore, the **utility of the forced upward titration** period is to gradually get patients to their randomized fixed dose, rather than trying to separate out the effects of time and dose. Similarly, the **utility of the withdrawal period** is to gradually get patients completely off their fixed dose rather than trying to separate out the effects of time and dose reduction.

Another potential **analysis issue** may be how best to deal with the problem of **two primary efficacy endpoints**. Should the overall Type I error be spread across each endpoint per Bonferroni or in an unequal manner to reflect one endpoint having greater weight than the other? Or should a bivariate analysis be performed? Or should the two endpoints be combined in some meaningful way and the result analyzed univariately?

In addition, since patients move to higher doses only if the clinician believes they are non-responsive, there is the question of whether the **FTART** trial provides meaningful and **interpretable dose response information**. At the end of the fixed dosing period, there will almost surely be subgroups of patients in both the drug and placebo groups who were titrated to a fixed dose of D_i, i = 1, ..., 5. However, a comparison of the drug group to the placebo group provides the only randomization-based inference. This comparison, based on data at the end of the fixed dosing period, would provide an inference as to the effectiveness of the drug at a dose equal to the average of the doses over the dosing period. Although it is a post-randomization stratified analysis, the comparison could be carried out by blocking on the fixed doses achieved.

Further, there is a desire to **combine results** across both the FTDRT and the FTART trials in a dose response sense. Randomization ensures a valid inference base for pair-wise comparisons of each dose to placebo from the FDTRT trial; but not so from the FTART trial. One way of combining results from both trials would be to estimate the effect of all dose groups combined (appropriately weighted) as compared to placebo from the FTDRT trial and combine this estimate with the estimate of the drug group compared to placebo from the FTART trial using meta-analysis methods such as the Cochran-Dersimonian-Laird procedure [12]. This would then provide an overall estimate of the effectiveness of the drug over the range of doses D_1, ..., D_5 or at the average of the doses -- simultaneously incorporating any heterogeneity across the trial estimates of effectiveness.

Efficacy analyses described are endpoint analyses; i.e. use the last observation available on patients in the fixed dosing period. As such the inference reflects the degree to which the drug reduces the NPA or reduces the severity of the overall panic condition as reflected by the CGIS beyond such reductions in the placebo group. Typically, there is little or no interest in the **duration of the effect**. Whether duration of effect can be assessed will depend on the length of T over the upward titration period and/or on the length of the fixed dosing period. One definitional way to consider duration of effect is to define response as 0 panic attacks or a pre-specified reduction in the NPA. Once response is observed, particularly during the fixed dosing period, duration of response could be estimated using survival data analysis methods. These results would be interesting clinically, but would be difficult to interpret statistically as the group of responders is a subset of all randomized patients. In addition, survival data analysis methods could be used to provide a valid inference between drug and placebo groups in terms of **time-to-response patterns**. Yet another measure that may be of clinical interest is **the proportion of the dosing period patients are in the response state**. As non-responders would have a value of zero for this measure, comparison of the drug and placebo groups would be based on all randomized patients.

Finally, **modifications to the withdrawal phase** of the trials could lead to better understanding of withdrawal effects as well as provide an inferential framework for conclusions. At the end of the fixed dosing period patients could be re-randomized to a variety of de-escalation regimens to better characterize withdrawal effects. Good clinical judgment will be required in defining such de-escalation schemes so as to minimize severe withdrawal effects from abrupt reduction of the fixed dose. It is suggested that patients be seen at the clinic at more frequently scheduled visits (and effects: NPA, CGIS, and any adverse experiences, recorded at such visits).

Acknowledgement

The author expresses thanks to Ms. Ruth Whitworth of the Karl E. Peace Center for Biostatistics, to Miss Lee Mitchell, Senior Secretary, Office of the Dean, Jiann-Ping Hsu College of Public Health, Georgia Southern University, and to the editorial staff of the PSA for their assistance in improving the manuscript for publication.

References

[1] Pary, R, Lewis, S: *"Identifying and treating patients with panic attacks."* American Family Physician; Sept., 1992.

[2] Grant BF, Hasin DS, Stinson FS, *et al*: *"The epidemiology of DSM-IV panic disorder and agoraphobia in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions."* J Clin Psychiatry 2006; 67:363–74.

[3] Code of Federal Regulations: *Sec.314.126 Adequate and well-controlled studies.* Title 21, Volume 5 [21CFR314.126]; U.S. Government Printing Office via GPO Access; Revised as of April 1, 2005, pp. 144-146.

[4] Rosenbaum JF, Moroz G, Bowden CL: *"Clonazepam in the treatment of panic disorder with or without agoraphobia: a dose-response study of efficacy, safety, and discontinuance."* J Clin Psychopharmacol 1997; 17:390-400.

[5] Davidson JR, Moroz G: "*Pivotal studies of clonazepam in panic disorder.*" Psychopharm Bull., 1998; 34:169-174.

[6] Moroz G; Rosenbaum JF: *"Efficacy, safety, and gradual discontinuation of clonazepam in panic disorder: a placebo-controlled, multi-center study using optimized dosages."* J Clin. Psychiatry, 60(9):604-12, 1999.

[7] American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 4Ed (DSM-IV). Washington, DC: APA, 1994

[8] Peace, K.E.; *"One-Sided or Two-Sided p-Values: Which Most Appropriately Address the Question of Drug Efficacy?";* J. Biopharm. Stat.; 1 (l): 133-38; 1991.

[9] Peace, K.E.; *"Sample Size Considerations in Clinical Trials Pre-Market Approval."* The Philippine Statistician; 55 (1, 2): 1-26; 2006.

[10] Peace, K.E.: *"Validity of Statistical Inferences from Medical Experiments."* Philippine Statistical Association Newsletter; 32 (2): 10-16; August, 2006.

[11] Mantel N, Haenszel W: *"Statistical aspects of the analysis of data from retrospective studies."* Journal of the National Cancer Institute,1959; 22:719-748.

[12] DerSimonian R, Laird N: *"Meta-Analysis in Clinical Trials."* Controlled Clin Trials, 1986; 7: 177-188.