Adaptive Randomization Study Design in Clinical Trials for Psychiatric Disorders

Roberto Gomeni[1]* and Navin Goyal[2]

[1]Pharmacometrica, Longcol, La Fouillade, France
[2]Clinical Pharmacology Modeling and Simulation, GlaxoSmithKline, King of Prussia, PA, USA

Abstract

Background: Uncontrolled placebo response has been one of the major culprits in the failure of randomized clinical trials for depression. A major drawback associated with the presence of high placebo response is the increased noise-to-signal ratio that in a majority of cases prevents the detection of treatment effect. The aim of this work was to propose an adaptive randomization study design based on band-pass filtering and evaluate this approach when compared to the traditional study designs.

Results: Clinical trial simulations demonstrated that an adaptive randomization approach always outperformed the conventional study designs in improving the signal-to-noise ratio. The improvement directly correlated with the level of placebo response and the variability in response across centers. The proposed strategy does not warrant any un-blinding of data.

Conclusions: The use of the adaptive randomization design provides a novel methodological approach for signal detection in clinical trials where placebo represents a known confounding factor. The improvement in signal detection was directly proportional to the level of placebo response, the degree of heterogeneity across recruitment centers with a reduced sample size as compared to the traditional study design. These findings support the use of the band-pass filtering approach in an adaptive randomization design as an efficient way to minimize the impact of uncontrolled placebo response and provide rational go/no-go decision criteria in the development of new medicines for psychiatric disorders.

Keywords: Adaptive randomization; Placebo response; Band-pass filtering

Introduction

A meta-analysis conducted on the US Food and Drug Administration database, including 12 approved antidepressant drugs, 74 randomized clinical trials (RCTs), and 12,564 patients, indicated that 49% of the clinical trials failed [1]. Converging findings indicate placebo response as the most relevant among the factors contributing to the failure of RCTs [2-4]. Furthermore, a recent meta-analysis showed a correlation between different levels of placebo response rate and clinical trial outcome in major depressive disorder [5]. The results of this meta-analysis suggest that the relative efficacy of the active drug compared to placebo in clinical trials for MDD is highly heterogeneous across studies with different placebo response rates, with a worse performance in showing a superiority of the drug versus placebo for studies with placebo response rates ≥ 30% and ≥ 40%, respectively, for monotherapy and adjunctive trials. The conclusion suggest that It is important to maintain placebo response rates below this critical threshold, since this is one of the most challenging obstacles for new treatment development in MDD. Several factors have been invoked to interpret and explain the level of placebo response, including management of patient expectation by the investigator, investigator bias about the efficacy of the new treatment, misdiagnosis, and regression to the mean. Confounding motivations, often implicit, may result in setting high expectations about the treatment outcome (placebo and active arms), with investigators desiring to heal the patient and patients wanting to please the investigators, leading to high placebo response. In addition, diagnostic misclassification could lead to the inclusion of subjects that do not have the psychiatric disorder under consideration, sometimes leading to spontaneous improvements that are indistinguishable from placebo response.

A major assumption in any multi-centered RCTs is that all recruitment centers perform similarly. The sample size, power or analysis plan of any multicenter RCT is based on the assumption that every recruitment center to which treatment (active or placebo) is applied generates interchangeable and homogeneous informations. The homogeneity across various recruitment centers is assured by the randomization process. Analyses of historical data demonstrate that such an assumption is unfounded and high heterogeneity across recruitment centers has been observed [6]. One of the possible reasons for such an observation is the complex interaction of expectations between the patients (hoping to be cured) and the investigators (hoping to discover novel pharmacological treatment for the disease condition e.g. depression). Each center in a multi-centered RCT may manage such expectations of trial outcomes differently. This often may lead to scenarios where centers record inflated improvements in disease status (both active and placebo). All these factors result in wide range of placebo effects. A formal analysis to evaluate the impact of non-informative centers on the outcomes of a clinical trial has been conducted [6]. In a typical study, an average of ~40% of the centers was classified as non-informative. This finding suggests that only 60% of the original study population is informative as regards the detection of potential clinical effects of active drugs. The signal of treatment effect (difference between the end-of-study HAMD-17 scores for active treatment and placebo) was ~80% higher in the informative center

data set when compared with the original data set, for both the active treatment arms.

Results from a disease-drug-trial modeling framework have demonstrated the utility of applying a band-pass filtering approach to improve signal detection in the research and development of novel drugs in psychiatric diseases [7]. A "window" or "band" of acceptable/plausible placebo response outcomes has been proposed to filter out centers generating implausible data based on sufficient clinical experience and historical data. We have also previously demonstrated that the treatment effect varies as a function of placebo response irrespective of the dose used in the clinical trial. Details of this approach have been discussed in literature [8].

Development of novel medications for treating psychiatric diseases has become highly risky and difficult due to limited resource availability. In the last few years several large pharmaceutical companies discontinued or severely restricted their central nervous system (CNS) drug development investments, due to the high cost, the long drug development duration, and the disproportionately low chances of success [9]. This has created an urgent need for novel clinical study designs. Such designs should be efficient and provide relevant information in making a well-informed go/no-go decision early on and focus on candidates with high probability of success. Adaptive study designs are gaining increasing popularity to this end as they allow (as the name suggests) adapting trial design based on emerging data [10]. This manuscript presents the use of band-pass filtering in an adaptive randomization study design framework. Such an approach may provide an efficient solution to control the placebo response issue. Virtual Phase IIb and III trials were simulated to compare the outcome of adaptive study with the traditional study design.

Methods and Data

Virtual, multicenter, placebo controlled Phase IIb trials were simulated to assess the outcomes for an antidepressant drug being developed for MDD. The model utilized in the simulations was developed and validated using data from five RCTs studying paroxetine and included a total of 1837 MDD patients from 124 recruitment centers [7]. The selection of the trials included in the model development was based on similarities in their key design factors, i.e. depression severity at baseline (HAMD ≥ 23), number of treatment arms (n=3), and the year of publication (2002-2004). The duration of these studies was between 8 to 12 weeks.

The clinical response (to either placebo or drug) was defined by the time-varying HAMD scores, considered the "standard" endpoint in MDD RCTs. The trajectory of this curve usually shows a nonlinear decrement from a high initial score (e.g., ~23) to lower values (e.g. ~10) associated with clinical remission, within 6-8 week of treatment, the typical time-lag to detect reliable clinical effect in MDD. The HAMD time-course in each treatment arm was analyzed using a mixed Weibull/linear equation:

$$f(t) = Ae^{-(t/td)^b} + h_{rec}t \qquad (1)$$

A, b, td and h_{rec} were the fixed effect parameters. A represents the baseline HAMD score, t_d is the time corresponding to 63.2% of the maximal change from baseline, b is the shape or sigmoidicity factor, and hrec is the remission rate. This model has been successfully applied for describing: a) the placebo response in 9 RCTs [6], b) the placebo response in 7 RCTs [11] and subsequently to describe either the placebo or the active drugs response in 5 RCTs [7]. Based on the good predictive performances of the model, the simulations conducted in

the present analyses are expected to generate outcomes of a new RCT that closely reflect real life data. The model parameters were estimated using NONMEM [12]. The random effects were assumed normally distributed for A and log-normally distributed for td, b and h_{rec} with a zero mean and variance Ω with a proportional residual error model. The mean placebo responses of each recruitment center were estimated by averaging the Bayesian post-hoc individual parameter estimates by center. Details of the population parameter estimates have been reported elsewhere [9].

The simulated trials were 8 weeks long. Virtual clinical trials were generated using a Monte Carlo simulation approach. This method is based on a stochastic model describing clinical effect over time in individual subjects. Outcomes are modeled as a function of subject characteristics, including drug, disease and placebo effect, any generic covariates, and random effect. Each simulation draws a new set of subjects from a virtual population based on a pre-defined model and parameter distributions. The model, covariates, and trial protocol provide the framework for simulating a range of possible outcomes for a trial.

The clinical trial simulation (CTS) framework is summarized in Figure 1. The aim of the proposed adaptive design is to detect the level of placebo response in each recruitment center during the patient accrual using un-blinded observations. Based on the result of the analysis, an adaptive randomization process is applied to minimize the number of patients randomized in the uninformative centers and to maximize the number of patients enrolled in the informative centers. The uninformative centers are the one with an implausible (excessively high or excessively low) level of placebo response. The proposed algorithm and approach allows for the classifying a center without breaking the blind. Two hundred double-blind, 8 week long RCTs with placebo and active treatment arms were simulated. These simulations were repeated for different levels of placebo response. As displayed in Figure 1, two different simulation approaches were applied. These approaches are discussed below in detail.

Conventional study design

This approach involved selecting and initiating inclusion of patients at all of the planned recruitment centers to achieve a targeted level of enrollment. In each simulated RCT, 8 subjects were enrolled in the placebo group and similar number on active treatment at each recruitment center. Each RCT was expected to recruit patients in 40 centers. This reflects the typical number of recruitment centers and subjects that are enrolled in RCTs for MDD. In each trial, the treatment effect (TE) was estimated as the mean difference between the baseline-corrected clinical scores in the active and in the placebo arm at the end of the study. No band-pass filtering was used in this approach.

Placebo response=HAMD score at baseline (on placebo)-HAMD score at week 8 (on placebo) (2)

Drug response=HAMD score at baseline (on drug)-HAMD score at week 8 (on drug) (3)

Treatment Effect (TE)=Drug response–Placebo response (4)

Adaptive randomization study design

The adaptive randomization study design was implemented as follows: the trial was initiated by starting recruitment at a limited number of recruitment centers. In the simulated RCTs, ten recruitment centers were used. It was assumed that each center could approximately enroll eight or more subjects in each arm. The simulated
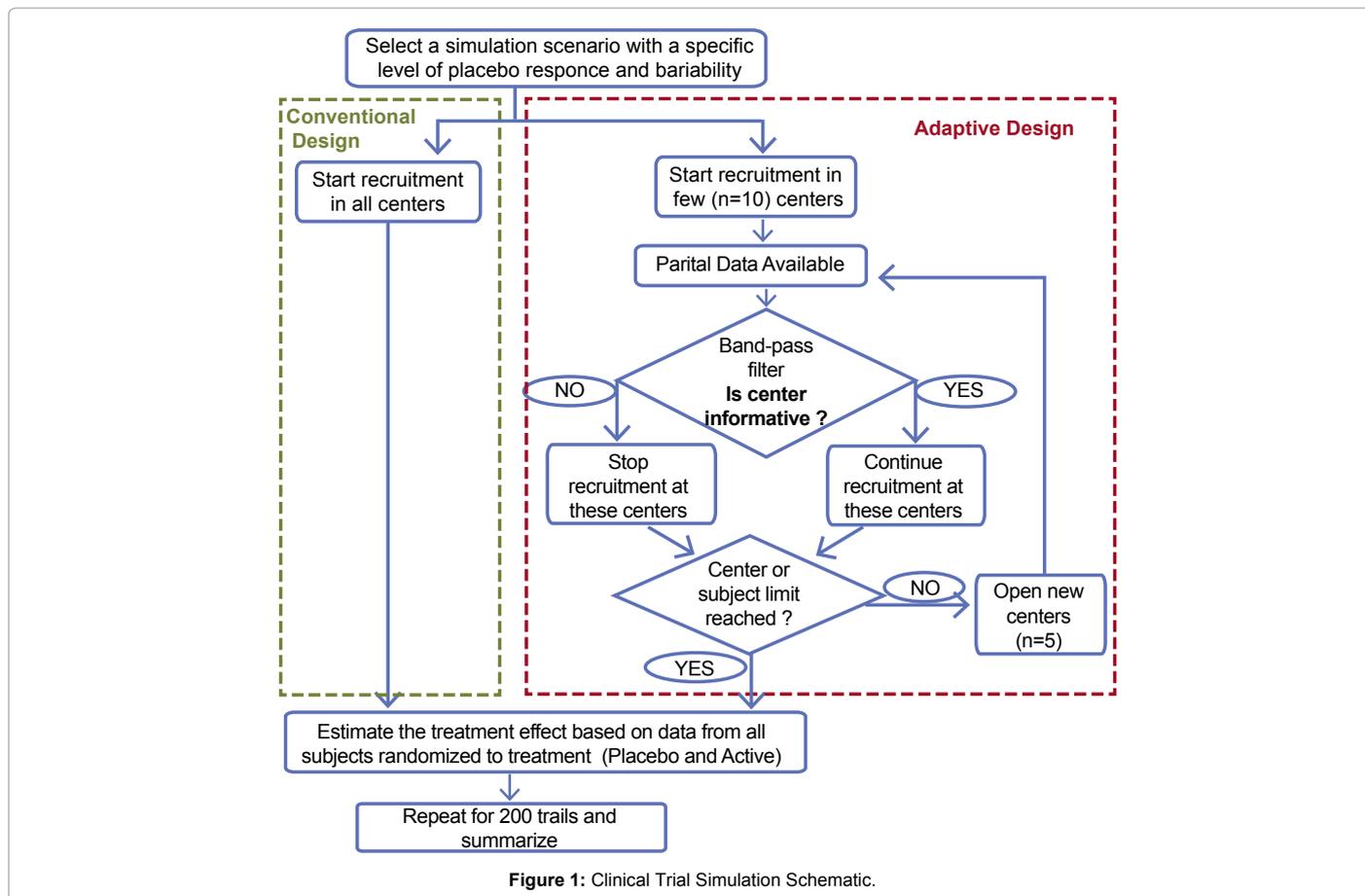
Figure 1: Clinical Trial Simulation Schematic.

study design assumed a 1:1 active to placebo allocation. A Bayesian modelling analysis was conducted to evaluate the minimum sample size to provide an acceptable precision of the typical placebo response of each center. This analysis indicated that a minimal sample of four subjects would be required [6]. Once data was available from at least four subjects each, on the placebo and the active treatment arm in a center, the un-blinded data in that center was analyzed to classify the center as informative or uninformative based on band-pass filtering. In this analysis, the lower and the upper bound of the filter were set at 11 and 20 points on the HAMD scale. The estimates of the cut-off values were based on clinical criteria for the definition of clinically relevant improvements in the disease condition given a specific baseline disease severity at enrollment. The partial data were analyzed using a non-linear mixed-effect model (Equation 1) and the estimates of the HAMD scores at study-end in each center were estimated. If collectively more than two-thirds of the subjects at a center had their HAMD scores at week 8 outside the 11-20 HAMD point range, the center was classified as uninformative.

The two-thirds rule was empirically selected among different evaluated strategies. This rule was based on practical considerations. If a drug is highly effective in improving the disease condition one can expect most subjects in active treatment arm to have their HAMD scores at week 8 below 11 points. However, there is a limited likelihood that more than two-thirds (67%) of the subjects at a center, which include subjects randomized to placebo (1:1) as well, achieve HAMD =11 that corresponds to a condition close to partial remission (HAMD =10), a rarely occurring improvement in MDD within this timeframe

[13]. This would mean that all subjects in active arm and at least one-third subjects in the placebo arm would demonstrate significant improvement in their disease status. Other options such 60% or 75% of subjects passing the band were also evaluated. This can be easily adjusted based on the disease area, baseline clinical scores, historical and expected placebo response. The upper cut-off limit of HAMD =20 at week 8 corresponded to a reduction of less than 10% over their baseline at recruitment. This threshold was based on the outcome of a meta-analysis on historically placebo response study, showing that a reduction, relative to baseline, of <10% (given the inclusion criterion of HAMD score ≥ 23), was highly unlikely [14]. Converse to the scenario of demonstrating high response with active and placebo treatments, this would reflect center with no improvement at all with either treatment. This would also be an indication of an uninformative center.

Once the center was classified as uninformative, further enrollment was discontinued at those centers and new centers were opened. During the CTS, five new centers were opened at a time. While patient recruitment was underway at new centers, recruitment also continued at the previously opened centers that were informative based on the band–pass filtering. Once data was available from at least four subjects on each arm at the newly opened centers, they were evaluated similarly using the HAMD 11 to 20 point window wherein recruitment was terminated at uninformative centers and continued at informative centers. The process was continued until either a pre-defined targeted number of centers or patient numbers at the centers were achieved. The trial was stopped when about 40 centers in total were opened for recruitment.

The TE with the adaptive approach was calculated by utilizing data from all subjects at all the centers that were randomized to placebo and active treatment. This includes subjects at informative and uninformative centers. However, since recruitment was terminated at some centers, based on band-pass filtering, the number of subjects available to estimate the TE in an adaptive design will always be ≤ number of subjects available with the conventional study design.

Evaluation

Three different levels of placebo effect (td=4, 4.5 and 6) were considered to evaluate the impact of the heterogeneity in placebo response in the different recruitment centers on the TE. The different levels of heterogeneity were simulated assuming different levels of between centers' variability. The TE improvement was calculated as the percentage improvement in TE (%TE) between adaptive and conventional study design over the conventional study design.

$$\%TE\ improvement = (TE_{adaptive} - TE_{conventional}) * 100 / TE_{conventional} \qquad (5)$$

Results

Parameters listed in Table 1 were employed in CTS to generate the individual HAMD profiles for subjects randomized to placebo and active treatment over eight weeks across different centers. The improvement in a subject's condition was estimated as the difference between the HAMD scores at week 8 and baseline. The mean difference in improvement across treatments was calculated as the TE. The %TE improvement was estimated as listed in equation 5. Outcome from such a model based adaptive randomization study design displayed in Figure 1 is discussed below.

The results from these CTS confirm the results of our previous work that the expected treatment effect is more reliably estimated if data from informative centers is used. There is significant improvement in the signal detection with the adaptive band pass-filtering technique. Combining the band pass filtering approach with an adaptive study design results improved signal-to-noise ratio and reduced placebo response rates. The adaptive approach was stopped once 40 centers were opened and involved terminating new patient recruitment at uninformative centers. Consequently, the number of subjects in the final analysis with the adaptive design is significantly lower as compared to the traditional study design.

High placebo effect

A high placebo effect that is almost comparable to the response at lower dose of the active treatment was evaluated. Scenarios evaluated variability between centers ranging from 35% to 70% in stepwise increments while the within center variability was kept constant at around 50%. It can be seen from the results displayed in Figure 2 that utilizing an adaptive study design leads to an improved detection of treatment effect as compared to the conventional study design. This improvement in signal detection with adaptive approach increases as the variability across the study centers increases as compared to the variability within a center. In the current CTS, there was 2-21% TE improvement over the range of variability evaluated. Such an improvement can often impact the go/no-go decisions in drug development. Approximately 30-35% fewer subjects would be recruited/needed with the adaptive approach.

Low and intermediate placebo effect

A low or intermediate placebo effect implies a lower impact of placebo response in differentiating or establishing the presence of benefit from the active treatment. While such scenario is highly likely to generate unbiased data, the adaptive study design will lead to even better quality data by reducing noise due to biased study centers. The results of the CTS demonstrate that an adaptive strategy could lead to 2-12% (low placebo effect) or 2-18% (intermediate placebo effect) TE improvement as heterogeneity across the centers increases. The adaptive approach would result in 22-35% fewer subjects recruited in the trial as compared to the traditional study design based in the placebo response rate and response variability across study centers.

	A (Baseline HAMD)	*b* (Sigmoidicity)	h_{rec} (Disease Progression)
Population Estimate	23.8	0.90	0.68
Between Subject Variability	5.41	0.49	0.45

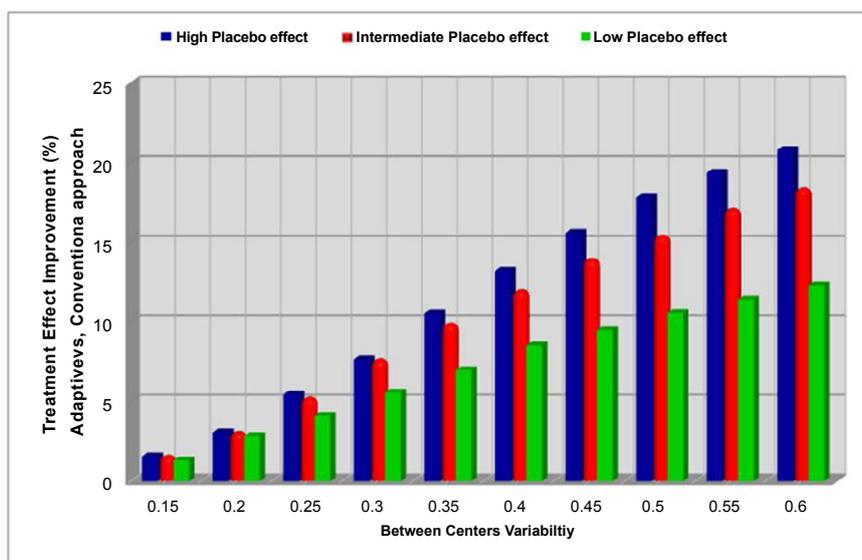**Table 1:** Parameter estimates used in simulations.

Figure 2: Treatment Effect improvement (%) with traditional and adaptive study designs.

Discussion

Understanding the effect of drug treatment versus the placebo effect remains a key challenging issue in the development of treatments for depression. Well established and effective anti-depressant drugs utilized as positive control in clinical trials for new treatments have also failed to differentiate from placebo in many trials [2]. It is well established that the heterogeneity across study centers in clinical trials for such treatments significantly impact the signal-to-noise ratio. To make matters worse, the level of noise generated is driven by the placebo response rate. Previous work has demonstrated the benefit of band-pass filtering to tackle this problem. However, this methodology was mainly considered as a post-hoc analysis approach to clean-up the data from the excessively high placebo response and, so far, for determining an unbiased signal of treatment effect. At variance of this approach, the current work propose to prospectively use the band-pass filtering methodology to early identify uninformative centers in an ongoing clinical trial and, based on this information, to implement an adaptive randomization scheme to stop the inclusion of patients in the uninformative centers and increase the inclusion of patients in the informative centers. Our approach consists in stopping the randomization of new patients in the center classified as un-informative and to progress (or expand) the randomization of new patients in the informative centers.

This approach is similar to the "play the winners" rule where a higher proportion of patients will be assigned to the centers less affected by the excessively high level of placebo response [15]. The "play the winners" method randomizes the next subject to the treatment group that was successful in the previous subject. At variance of this method, our approach proposes to expand the enrollment in the informative centers and stop the enrollment in the non-informative centers. In that sense the center is the winner. However, we do not "replace" the non-informative center. All the data already generated at the non-informative center is utilized in the final analysis.

The results from the CTS demonstrate the benefit of the adaptive randomization study design using the band-pass filtering approach. The benefit of the adaptive approach increases with the level of placebo response and the heterogeneity in response across the study centers. The adaptive process aims to stop the randomization of new patients in the uninformative centers and to increase the randomization of new patients in the informative centers. This process is aimed to limit the impact of data collected in uninformative centers and thus increase the overall probability of detecting a 'true' signal of treatment effect.

As recently shown, each individual recruitment center's efficiency in measuring actual clinical response is critically important for the overall success of a multicenter RCT [7]. However, even in centers with proven logistical capacity and adherence to protocol, difficulties in detecting TE are commonly observed. High levels of placebo response were observed in a substantial percentage of the centers within multicenter RCTs [6]. Several factors have been implicated in determining the level of placebo response, including management of patient expectation by the investigator, investigator bias about the efficacy of the new treatment, misdiagnosis, and regression to the mean. These factors are difficult to assess, and no methodologies are currently available to define a center's performance on the basis of the compounded contribution of each of these factors. The typical way to address this issue is to (i) select centers on the basis of the track record and (ii) provide awareness sessions and good clinical practice training to the investigators at the beginning of every multicenter RCT. Despite these efforts, heterogeneity among the performance levels of the centers remains relevant, calling for practical solutions [16]. In fact, clinical study findings repeatedly indicate that a center's performance is inconsistent over time and that awareness sessions do not have as much impact as expected [17].

There are other significant advantages of implementing an adaptive randomization design with the band-pass filtering approach. The number of subjects needed in such a study design is almost always less than that required with a conventional study design. The power achieved with a traditional study design can be obtained with fewer subjects using the adaptive study design discussed in this exercise. This is achieved by reducing the noise in the data while also improving the signal detection dependent on placebo response. As seen from the CTS, 20-35% fewer subjects were enrolled in the trial with the adaptive approach as compared to the traditional design. At the same time, there was increased %TE improvement across the placebo response level.

The choice of cut-off points for band-pass filtering in the CTS was based on prior information about the disease area, historical data and practical expectations around the placebo effects given the baseline disease severity inclusion/exclusion criteria. As such this can be easily adjusted/modified. While data from all subjects randomized to any treatment was included in the final analysis, the adaptive randomization design and subsequent analyses can very well be a-priori defined to exclude data from the uninformative centers. One can also reassess the previously defined informative or uninformative centers prior to final analysis based on latest data accrued. The number of data cuts to determine whether a center is informative or not can be further adapted on the basis of the number of study centers, number of subjects expected to be recruited at the center, recruitment rate and similar criteria.

There are certain assumptions/limitations in the current CTS. A 1:1 placebo to active treatment allocation would mean that while evaluating a center there are equal number of subjects on placebo and active treatment. The upper and lower limit for band-pass filtering can be adjusted accordingly if the randomization is different from 1:1. The limits could be adjusted to provide a conservative or lax filter. Alternatively, the proportion of subjects in a center that meet the cut off criteria can be altered accordingly. This approach is highly flexible to accommodate different disease areas and conditions. The proposed methodology is a simulation exercise and the results of the simulations need to be confirmed by real data generated in a real RCT. The lack of supportive experimental data represents the main limitation of the methodology. This approach focuses on the control of noise or bias generated by recruitment centers only and does not account for noise generated in the data by other factors. In addition missing data (or dropout rate) is not factored in the simulation model. However, this limitation can be easily resolved by applying a model recently proposed to jointly analyze longitudinal scores described by a Weibull/linear model and dropout events [18]. Finally, the implementation of this approach will result in an increased time for completing a trial as well as in an increased logistical complexity for conducting a trial. This may result in the overall inflation of the cost for a trial although the approach in general should decrease overall drug development cost by providing data with better signal to noise ratio and a reduced risk of having failed or uninformative trials.

Conclusion

Clinical trial simulation demonstrated the benefit of employing the proposed adaptive randomization approach in evaluating informative recruitment centers for patient allocation as compared to the traditional study design. The use of the adaptive randomization approach

provided a novel methodological approach for signal detection in clinical trials where placebo effect represents a known confounding factor. The improvement in signal detection was directly proportional to the level of placebo response, the degree of heterogeneity across recruitment centers with a reduced sample size as compared to the traditional study design. These findings support the use of the band-pass filtering approach in an adaptive randomization design as an efficient way to minimize the impact of uncontrolled placebo response and provide rational go/no-go decision criteria in the development of new medicines for psychiatric disorders.

References

1. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R (2008) Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med 358: 252-260.

2. Khan A, Detke M, Khan SR, Mallinckrodt C (2003) Placebo response and antidepressant clinical trial outcome. J Nerv Ment Dis 191: 211-218.

3. Kaptchuk TJ, Kelley JM, Deykin A, Wayne PM, Lasagna LC, et al. (2008) Do "placebo responders" exist? Contemp Clin Trials 29: 587-595.

4. Papakostas GI, Fava M (2009) Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. Eur Neuropsychopharmacol 19: 34-40.

5. Iovieno N, Papakostas GI (2012) Correlation between different levels of placebo response rate and clinical trial outcome in major depressive disorder: a meta-analysis. J Clin Psychiatry 73: 1300-1306.

6. Merlo-Pich E, Gomeni R (2008) Model-based approach and signal detection theory to evaluate the performance of recruitment centers in clinical trials with antidepressant drugs. Clin Pharmacol Ther 84: 378-384.

7. Merlo-Pich E, Alexander RC, Fava M, Gomeni R (2010) A new population-enrichment strategy to improve efficiency of placebo-controlled clinical trials of antidepressant drugs. Clin Pharmacol Ther 88: 634-642.

8. Goyal N, Gomeni R (2013) Exposure-response modeling of anti-depressant treatments: the confounding role of placebo effect. J Pharmacokinet Pharmacodyn 40: 389-399.

9. Kaitin KI, Milne CP (2011) A Dearth of New Meds: Drugs to treat neuropsychiatric disorders have become too risky for Big Pharma. Scientific American 305: 16.

10. Merlo-Pich E, Bettica P, Gomeni R (2009) Bayesian Monitoring and Bootstrap Trial Simulation: A New Paradigm to Implement Adaptive Trial Design for Testing Antidepressant Drugs. The Open Psychiatry Journal. 3: 20-32.

11. Gomeni R, Merlo-Pich E (2007) Bayesian modelling and ROC analysis to predict placebo responders using clinical score measured in the initial weeks of treatment in depression trials. Br J Clin Pharmacol 63: 595-613.

12. Beal SL, Sheiner LB, Boeckmann AJ (1989-2014) NONMEM Users Guides. ICON Development Solutions, Elicott City, Maryland.

13. Rush AJ, Warden D, Wisniewski SR, Fava M, Trivedi MH, et al. (2009) STAR*D: revising conventional wisdom. CNS Drugs 23: 627-647.

14. Khan A, Detke M, Khan SR, Mallinckrodt C (2003) Placebo response and antidepressant clinical trial outcome. J Nerv Ment Dis 191: 211-218.

15. Rosenberger WF, Huc F (2004) Maximizing power and minimizing treatment failures in clinical trials. Clin Trials 1: 141-147.

16. Kobak KA, Feiger AD, Lipsitz JD (2005) Interview quality and signal detection in clinical trials. Am J Psychiatry 162: 628.

17. Rickels K, Robinson DS (2007) Why do clinical trials fail? J Clin Psychopharmacol 27: 420-421.

18. Gomeni R, Lavergne A, Merlo-Pich E (2009) Modelling placebo response in depression trials using a longitudinal model with informative dropout. Eur J Pharm Sci 36: 4-10.