Neuropsychopharmacology (2015) 40, 2588–2595

www.neuropsychopharmacology.org

A Novel Methodology to Estimate the Treatment Effect in Presence of Highly Variable Placebo Response

Roberto Gomeni*[,1], Navin Goyal[2], Françoise Bressolle[1] and Maurizio Fava[3]

[1]R&D Department, Pharmacometrica, Longcol, La Fouillade, France; [2]Clinical Pharmacology Modeling and Simulation Department, GlaxoSmithKline, King of Prussia, PA, USA; [3]Psychiatry Department, Massachusetts General Hospital, Boston, MA, USA

One of the main reasons for the inefficiency of multicenter randomized clinical trials (RCTs) in depression is the excessively high level of placebo response. The aim of this work was to propose a novel methodology to analyze RCTs based on the assumption that centers with high placebo response are less informative than the other centers for estimating the 'true' treatment effect (TE). A linear mixed-effect modeling approach for repeated measures (MMRM) was used as a reference approach. The new method for estimating TE was based on a nonlinear longitudinal modeling of clinical scores (NLMMRM). NLMMRM estimates TE by associating a weighting factor to the data collected in each center. The weight was defined by the posterior probability of detecting a clinically relevant difference between active treatment and placebo at that center. Data from five RCTs in depression were used to compare the performance of MMRM with NLMMRM. The results of the analyses showed an average improvement of ~ 15% in the TE estimated with NLMMRM when the center effect was included in the analyses. Opposite results were observed with MMRM: TE estimate was reduced by ~ 4% when the center effect was considered as covariate in the analysis. The novel NLMMRM approach provides a tool for controlling the confounding effect of high placebo response, to increase signal detection and to provide a more reliable estimate of the 'true' TE by controlling false negative results associated with excessively high placebo response.
Neuropsychopharmacology (2015) **40,** 2588–2595; doi:10.1038/npp.2015.105; published online 20 May 2015

INTRODUCTION

The very-high frequency of failed and negative randomized clinical trials (RCTs) has been recognized as a critical issue for the clinical development of novel medications, especially in central nervous system therapeutic area (Stein *et al*, 2006; Alexopoulos *et al*, 2007). One of the most common reasons for study failure is the unexpected and excessively high level of placebo response (Khan *et al*, 2003). The interactions between patients included in a trial and investigator at recruitment centers have been recognized as one of the major source of placebo response (Allan and Siegel, 2002; Trivedi and Rush, 1994; Benedetti *et al.* 2005).

The level of placebo response has been shown to strongly impact the probability of detecting active treatment superiority (Iovieno and Papakostas, 2012). Furthermore, growing evidence indicates that placebo response in antidepressant trials has been gradually increasing over time (Walsh *et al*, 2002). These findings indicate that there is an urgent need for exploring, evaluating and implementing novel study designs and data analyses methodologies to tackle the uncontrolled and time-varying level of placebo response in these trials. Critical thinking and novel approaches are essential to improve the overall efficiency of clinical trials.

Different strategies have been proposed to account for a higher-than-anticipated placebo response due to the difficulties in precisely identifying placebo responders at inclusion time, predicting placebo response rates, and reducing the level of placebo response. These strategies included study designs with increased sample size, increased symptom severity at baseline, innovative study designs, enhanced inter-rater reliability programs, surveillance of within-study data to identify measurement error, site-independent subject validation to minimize site-biases, and enhanced patient education to minimize expectancy effects (Fava *et al,* 2003; Kobak *et al,* 2007; Targum *et al,* 2008; Targum *et al,* 2012).

Among the different approaches for optimizing RCTs, the sequential parallel comparison design (SPCD) was proposed to reduce both placebo response and sample size (Fava *et al,* 2003; Papakostas *et al,* 2014). The SPCD involves two double-blind stages of treatment, with stage 2 commencing immediately at the conclusion of stage 1. Typically, only subjects identified as not responding to placebo during stage 1 are included in the efficacy analysis of stage 2. The relevant data from the two phases are pooled to compute an overall *P*-value. Because of this pooling approach, use of SPCD with a specified sample size can provide a reduction in *P*-value in comparison with conventional parallel designs (Heger 2013; Fava *et al*, 2012). A number of RCTs have been conducted based on the SPCD design and the results confirmed that

*Correspondence: Professor R Gomeni, R&D, Pharmacometrica, Lieu-dit Longcol, La Fouillade, 12270, France, Tel: +33 760451976, Fax: +33 983233188, E-mail: roberto.gomeni@pharmacometrica.com
Received 7 January 2015; revised 17 March 2015; accepted 8 April 2015; accepted article preview online 21 April 2015

Table 1 Details of the Five RCTs Included in the Meta-Analysis. CR and IR Denote Controlled and Immediate Release Formulations Respectively

Study	Nb of Centers	Nb of Patients	Arm 1		Arm 2	
			Treatment	Dose	Treatment	Dose
448	19	299	Parox_IR	20–50 mg -flex	Parox_CR	25–62.5 mg -flex
449	20	333	Parox_IR	20–50 mg -flex	Parox_CR	25–62.5 mg -flex
487	26	319	Parox_IR	10–40 mg -flex	Parox_CR	12.5–50 mg -flex
810	38	489	Parox_CR	12.5 mg-fix	Parox_CR	25 mg-fix
874	21	397	Parox_CR	12.5 mg-fix	Parox_CR	25 mg-fix

SPCD provides a large reduction in *P*-value with an increase in power (Papakostas *et al,* 2014; Fava *et al,* 2012).

In an attempt to reduce placebo response, a double blinded placebo lead-in phase was proposed to detect the level of placebo response prior to the randomization of patients to the treatment arms. A meta-analysis conducted on the FDA database (86 RCTs in major depressive disorders (MDD) and 34 RCTs in schizophrenia) indicated that the placebo lead-in design was an efficient approach in reducing the number of placebo responders and thus reducing the average response of placebo (Chen *et al*, 2011). However, an analysis of 75 RCTs conducted from 1981 to 2000 in MDD patients showed that the average placebo response rate of studies that used a placebo lead-in did not differ significantly from that of studies that did not use a placebo lead-in (Walsh *et al,* 2002). These results confirm the findings of a meta-analysis previously conducted on 101 RCTs (Trivedi and Rush, 1994).

Recently, a band-pass filter analysis strategy was proposed as a way to control the excessively high/low placebo response rates and thereby enhance signal detection (Merlo-Pich *et al,* 2010; Gomeni 2014; Gomeni and Merlo-Pich, 2007; Gomeni and Merlo-Pich, 2012; Merlo-Pich and Gomeni, 2008). This strategy asserts that optimization of the signal-to-noise ratio can be achieved by prospectively identifying the cut-off values located at the high and low ends of the placebo response distribution curve and subsequently filtering-out the data of the centers that fall outside these boundaries (Khan *et al*, 2007; Kilts *et al,* 2009). This methodology has been applied to re-analyze data from negative antidepressant RCTs. The results of this analysis showed an important increase in signal (Targum *et al*, 2014).

The aim of this work was to propose a novel methodology to analyze the outcomes of an RCT using the center-specific level of placebo response as weighting factor to estimate the treatment effect. Data generated in centers with high placebo response will be considered as less informative than data generated in the other centers to estimate the 'true' treatment effect. This can be defined as the effect estimated in absence of any confounding noisy factor such as the placebo response. In contrast to the band-pass filtering approach, the novel methodology will include in the analysis all data of the patients randomized to the trial. This approach is based on a nonlinear longitudinal modeling of the clinical scores (NLMMRM), a natural extension of the linear mixed-effect modeling approach for repeated measures (MMRM) generally used for analyzing longitudinal data.

MATERIALS AND METHODS

Data

Data from five RCTs were used in the analyses. The same data were previously analyzed for the assessment of the band-pass filter approach as a novel population enrichment strategy (Merlo-Pich *et al,* 2010). Data were derived from GSK clinical databases (GSK clinical trial register [http://ctr.gsk.co.uk/medicinelist.asp]). The five trials were selected based on their similarities in key design factors, ie, depression severity at baseline (HAMD $\geqslant$ 23), number of treatment (TMT) arms (TMT = 0 for placebo and TMT = 1 and 2 for the low and high dose), and year of publications (2002–2004). Details of the five RCTs are shown in Table 1.

Clinical Response Model

The clinical response (to either placebo or drug) was defined by the time-varying HAMD scores, considered the 'standard' endpoint in MDD RCTs (Hedlund and Vieweg, 1979). The trajectory of this clinical score usually shows a nonlinear decrement from a high initial score (eg, ~ 23) to lower values (eg ~ 10) associated with clinical remission, within 6–8 week of treatment, the typical time-lag to detect reliable clinical effect in MDD (Nierenberg and Wright, 1999). In each of the five studies, the HAMD time-course in the three treatment arms were independently analyzed using a mixed Weibull/linear Equation 1:

$$f(t) = A\mathrm{e}^{-(t/td)^b} + h_{\mathrm{rec}}t + eps \qquad (1)$$

A, *b*, *td*, and h_{rec} were the fixed effect parameters and *eps* the residual error. 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 h_{rec} is the remission rate. This model has been successfully applied to describe the placebo response in different RCTs (Gomeni and Merlo-Pich, 2007). The model parameters were estimated using the nonlinear mixed effect modeling approach (FOCE-I) implemented in the NONMEM software (Beal *et al,* 2009). The random effects were assumed normally distributed for A and log-normally distributed for *td*, *b* and h_{rec} with mean = 0 and variance = Ω with a proportional residual error model. The random effect for the baseline '*A*' was assumed normally distributed because the inclusion criteria in the studies limited the values of the HAMD score at the inclusion within a predefined range of values. Therefore, the normality assumption was

considered appropriate as we were not expecting extreme values of HAMD at baseline. At variance of this assumption, all the other parameters (including '*td*') were considered free to assume any possible positive value. As a consequence, these parameters were assumed to follow a log-normal distribution.

The mean placebo responses of each recruitment center were estimated by averaging the Bayesian *post-hoc* individual parameter estimates at those centers.

Estimate Treatment Effect Using MMRM

The treatment effect (TE) was defined as the difference in the baseline adjusted clinical score between the active drug and placebo arms at the end of the study.

The conventional approach applied for the evaluation of TE was based on the linear mixed-effects modeling approach for repeated measures (MMRM) (Mallinckrodt *et al,* 2008). This analysis was implemented in SAS (SAS Institute Inc, 2010) using an unstructured covariance matrix, time as a classification variable, and baseline measurement as a covariate, baseline x time interaction, and treatment × time interaction with a significance level of $\alpha = 0.05$ to establish the significance of the treatment effect. Two analyses were conducted using the MMRM approach: with and without the inclusion of the study center as a covariate in the model and the results of these analyses were used as a reference.

Estimate Treatment Effect Using NLMMRM

One of the main reasons for the inefficiency of multicenter RCTs is the uncontrolled and excessively high level of placebo response. Such a placebo response can be considered as noise that impairs the signal-to-noise ratio. To overcome this issue, a novel data analysis approach is proposed. The basic assumption of this methodology is that centers with high placebo response are less informative than the others for estimating the 'true' treatment effect. As a consequence the weight of the information generated in that center should be lower than the weight of the information generated in the centers with low placebo response.

The level of information associated with each recruitment center was defined by the performance level of that center (P) and the weighting factor for that center was assumed to be inversely proportional to P (weight $= 1/P$). The performance of a recruitment center was defined as the posterior probability of detecting a signal of a TE > 3 units in the HAMD clinical score and was estimated using a methodology previously developed (Merlo-Pich and Gomeni, 2008). The probability (P) of detecting a TE > 3 was estimated by the logistic model (Equations 2,3):

$$\lambda = \theta_1 + \theta_2 \cdot \mathrm{HAMD}_{17}(Baseline) + \theta_3 \cdot \mathrm{HAMD}_{17}(StudyEnd) \quad (2)$$

$$P = \frac{e^{\lambda}}{1 + e^{\lambda}} \quad (3)$$

with: $\theta_1 = 2.310$, $\theta_2 = 0.291$ and $\theta_3 = -0.278$.

The new method for estimating TE was based on the assumption that the HAMD longitudinal scores can be described by a nonlinear model and the parameters of this model can be estimated using a nonlinear mixed-effect modeling approach. This novel method will be referred to as nonlinear mixed-effect modeling (NLMMRM).

The longitudinal time-course of the HAMD scores in the three treatment arms (TMT $= 0$ for placebo and TMT $= 1$ and 2 for the low and high dose) were defined by the Equations 4,5,6:

$$F_0(t, TMT = 0) = A_0 e^{-(t/td_0)^{b_0}} + h_{\mathrm{rec}0} t + eps \quad (4)$$

$$F_1(t, TMT = 1) = A_1 e^{-(t/td_1)^{b_1}} + h_{\mathrm{rec}1} t + eps \quad (5)$$

$$F_2(t, TMT = 2) = A_2 e^{-(t/td_2)^{b_2}} + h_{\mathrm{rec}2} t + eps \quad (6)$$

Each equation was re-parameterized as a function of the HAMD score at week 8 (F8):

$$F8_i = A_i e^{-(8/td_i)^{b_i}} + h_{\mathrm{rec}i} 8 \quad (7)$$

and

$$h_{\mathrm{rec}}i = \frac{F8_i - A_i e^{-(8/td_i)^{b_i}}}{8} \quad (8)$$

The change from baseline of the HAMD score at week 8 in the placebo arm was defined by Equation 9:

$$\Delta p = F8_{\mathrm{i}} - A_{\mathrm{i}} \quad (9)$$

Using this value, the equations describing the treatment in the first arm (TMT $= 1$) and in the second arm (TMT $= 2$) were re-parameterized in order to include in the model a parameter defining TE_1 (treatment effect of the first arm) and TE_2 (treatment effect of the second arm):

$$TE_{1,2} = \Delta p - \left(F8_{1,2} - A_{1,2}\right) \quad (10)$$

The residual error model ('*eps*') was defined as the sum of two components:

$$eps = \varepsilon + W_{\mathrm{j}} \quad (11)$$

where *eps* is a random variable with means $= 0$ and variance equal to the sum of ε (an additive error component) and W_{j} is a center-specific weighting factor defined as $1/p_{\mathrm{j}}$ (p_{j} = level of the performance of center$_{\mathrm{j}}$). According to this approach, all data belonging to center j were affected by a weight W_{j}. The centers with higher W (reflecting higher uncertainty in the measurements at that center) contributed less to the estimation of the model parameters; consequently, the model parameters were predominantly influenced by the data in the informative centers (with a lower W).

For each one of the five RCTs selected for the analysis, the NLMMRM was then implemented in five stages:

1. Only the placebo data from each center were analyzed and the model predicted individual HAMD trajectories were used to derive the typical value of the placebo response at baseline and at end of the study in each recruitment center.
2. The performance of each recruitment center was estimated using the logistic model previously defined in Equations 2,3.

3. A new analysis data set was constructed by merging the original data set with the probability values estimated in step 2.
4. An initial NLMMRM analysis was conducted on the three-arm data without including the center performance as a weighting factor.
5. The final NLMMRM analysis was conducted on the three-arm data using the center performance as a weighting factor.

In the initial setting of the model, all the model parameters, including the *eps* term, were estimated by minimizing a likelihood function. In this framework, the final value of *eps* was expected to provide an estimate of the residual error. As one component of *eps* (W_j) was fixed, the estimation procedure was unable to provide the 'true' residual error estimate. To overcome this problem, the ε term was fixed to a 'reasonable' value of the residual error. This value was set equal to the residual error estimated in the analysis of the placebo data alone (step 1 in the analysis) assuming that the same residual error was affecting the three treatment arms in the final step of the analysis (step 5).

The statistical evaluation of TE resulting from the NLMMRM analysis was done using the Wald test (*Wald*) (Harrell, 2001). According to this approach, the maximum likelihood estimate of $TE_{2,3}$ were compared to zero, assuming that TE_i is χ^2 distributed with one degree of freedom. In the univariate case, the Wald statistic is defined by:

$$Wald^2 = \frac{(TE_i)^2}{var(TE_i)} \qquad (12)$$

The values $var(TE_i)$ were provided by NONMEM once the maximum likelihood convergence was achieved.

RESULTS

The summary results of the MMRM and the NLMMRM analyses conducted across the five RTCs are presented in Table 2. The TE estimates were obtained either by independently fitting the HAMD scores in each study to a mixed Weibull/linear equation using the NLMMRM approach or by analyzing the HAMD changes from baseline using the MMRM approach implemented in the SAS PROC MIXED procedure. The assessment of the statistical relevance of the estimated TE_i was done by using the Wald test for the NLMMRM approach and by using the LSMEANS option in SAS.

The results of the analyses of the five RCTs provided consistent results. The results of the MMRM analyses indicate that the inclusion of the study center as covariate in the model leaves unchanged or decreases the estimated TE. A complete opposite trend is observed with the NLMMRM analysis. In this case, the inclusion of the center-specific level of placebo response as a weighting factor had the net effect to increase the estimated TE value in all the studies. These findings are graphically presented in Figure 1.

Figure 2 displays the mean % change in signal detection (TE value) across all the five RCTs selected when the center effect is included in the analysis.

The average improvement in the TE estimated with NLMMRM associated with the inclusion of the center effect in the analysis was on average of 17% in the first arm (low doses) and of 13% in the second arm (higher doses). Opposite results were observed with MMRM: TE estimate was reduced by 3.8% at low doses and by 4.1% at the higher doses when the center effect was considered.

These results represent a critical finding since the progression of a new drug to late phase development stages often depend on the outcome of proof-of-concept studies initially conducted in phase II. Domino effect could then include selection of wrong doses (often higher doses

Table 2 Summary Tables of Results Obtained with the MMRM and the NLMMRM Approaches With and Without the Center Effect

Study	MMRM analysis				NLMMRM analysis			
	Arm1-Pla	*P*	Arm2-Pla	*P*	Arm1-Pla	*P*	Arm2-Pla	*P*
No center effect								
448	1	0.316	2.14	0.031	0.85	0.219	1.75	0.068
449	2.75	0.004	3.28	0.001	2.58	0.012	2.81	0.009
487	1.63	0.056	2.27	0.008	1.34	0.039	1.26	0.085
810	1.22	0.156	2.9	0.001	2.26	0.01	4.21	0.001
874	1.32	0.086	2.5	0.001	0.91	0.001	2.04	0.001
With center effect								
448	1	0.314	2.14	0.029	1.97	0.068	2.46	0.019
449	2.74	0.004	3.13	0.001	2.39	0.028	3.35	0.001
487	1.58	0.049	2.23	0.006	2.63	0.005	2.78	0.002
810	1.03	0.221	2.52	0.004	2.73	0.001	4.39	0.001
874	1.32	0.085	2.47	0.001	1.67	0.023	2.69	0.001

Arm1.2-Pla represents the TE for the two treatment arms and *P* the associated probability level.

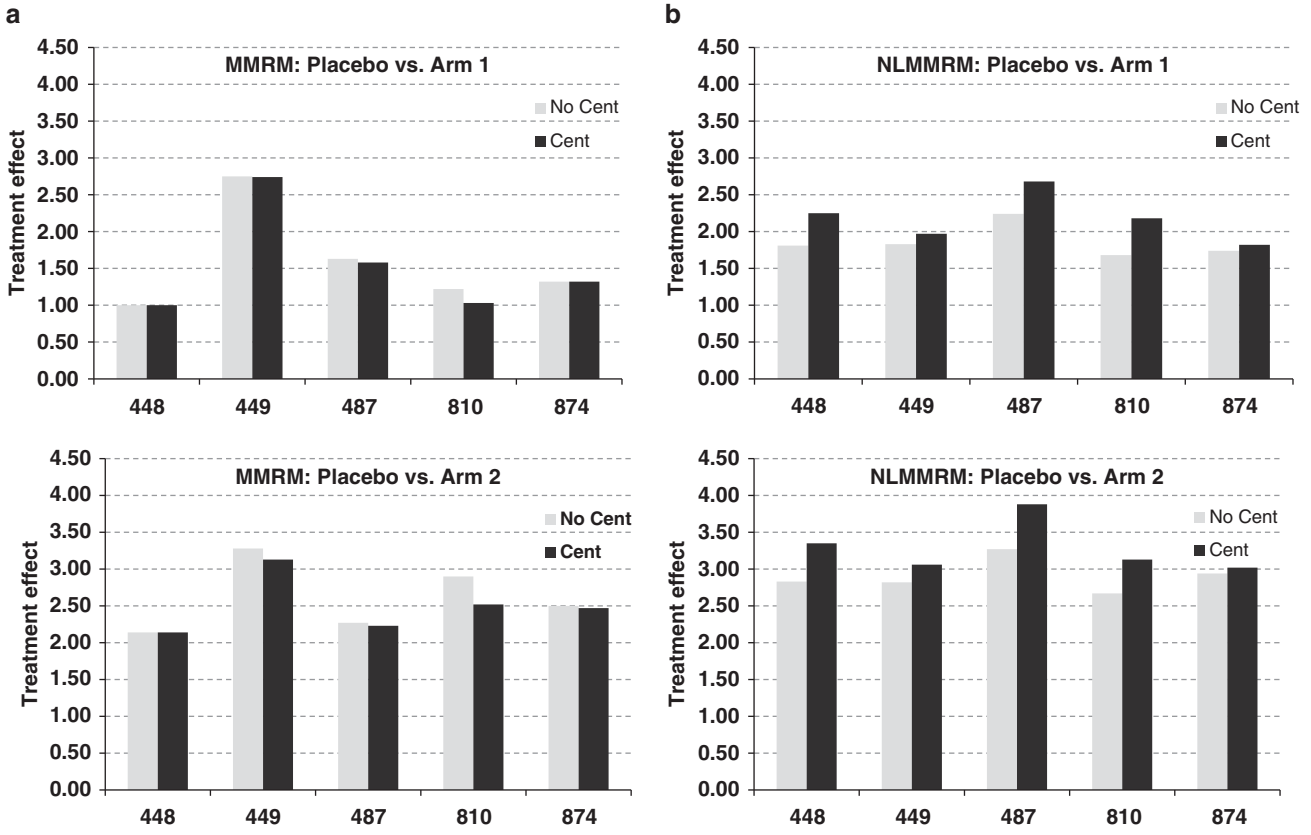

Figure 1 Treatment effect estimates using: (a) the linear mixed-effects model for repeated measures with and without the inclusion of center as a covariate in the model (light and dark gray bars), (b) the nonlinear mixed-effects model for repeated measures without and with the inclusion of the center performance as a weighting factor in the analysis (light and dark gray bars).

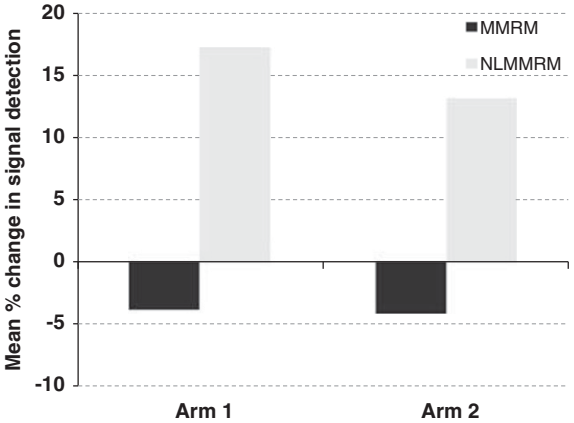

Figure 2 Mean % change in signal detection (treatment effect value) across the five randomized clinical trials considered with the modeling approach for repeated measures and the nonlinear longitudinal modeling of clinical scores approaches when the center effect is included in the analysis.

resulting in safety issues), dosing frequency or final decision to discontinue the clinical development of the drug.

The rationale for including covariates in a model is to increase the chance to find a statistical significant TE by explaining and reducing part of the variance of TE and therefore increasing the signal-to-noise ratio for TE. In the MMRM analyses, the center effect was always retained as statistical significant covariate however, the TE estimated with the center as covariate was lower than that estimated without the center as covariate. One of the possible reasons for this outcome is the violation of the assumptions required by MMRM to account for covariates.

MMRM assumes that the variability in the response variable (TE) can be explained by a linear combination of various constant levels corresponding to different combinations of the factors and/or a linear dependence on the values of the covariate(s). In all cases, the residual variations from such a hypothetical model are assumed to be independent normal deviates with constant variance. In addition, it is assumed that each level of the factor/covariate should be normally distributed with equal variance. This assumption would require that the TE should be equally distributed with equal variance in each center.

This assumption was strongly violated in each trial evaluated. As an example, the heterogeneous distribution of TE by center in the two treatment arms is shown in Figure 3 for the study 810. When the response in a MMRM model is observed under possibly nonparametric effects of a confounding covariate, a direct application of MMRM model may lead to biased estimates of the regression relationship of interest (Nguyen *et al*, 2008).

At variance of this limitation of the MMRM approach, the methodology used in the NLMMRM analysis for accounting for the center effect fully satisfies the modeling assumption.

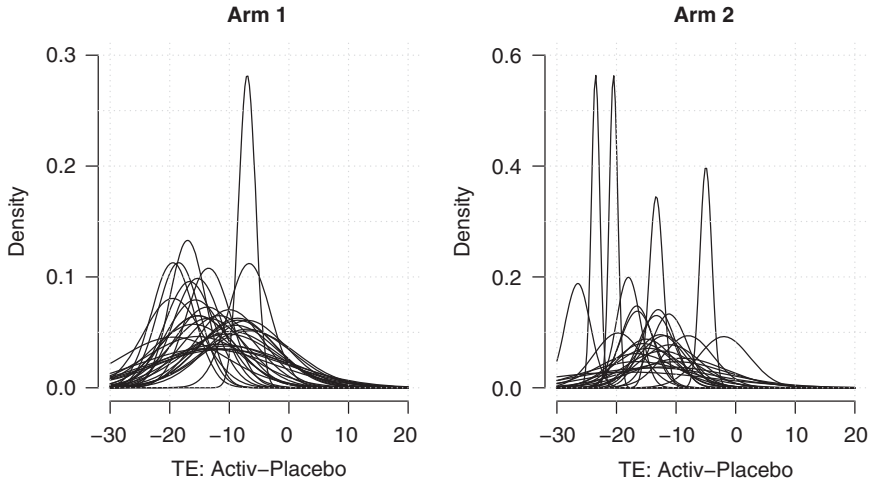

Figure 3 Probability density functions (computed assuming independent and normally distributed values) of treatment effect in each recruitment center for the study 810 by treatment arm.

In this approach, the center effect was assumed to be a component of the normally distributed residual error variance that, and in this way, was assumed to vary center by center.

DISCUSSION

Development of pharmacotherapy for CNS disorders has become increasingly challenging due to the enormous failure rates in their Phase IIB ad Phase III trials with poor separation of treatment related efficacy from placebo. Often, well established marketed treatments used as positive control also fail to distinguish from placebo.

The present paper presents a novel methodology to estimate the TE using a nonlinear mixed-effects model for repeated measures approach. NLMMRM can be considered as the natural generalization of the likelihood-based mixed-effects model for repeated measures (MMRM) approach that is today recognized as the most efficient and reliable method for conducting the primary analysis of continuous end points in longitudinal clinical trials. However, the analyses demonstrated that NLMMRM provides a more powerful methodology than MMRM for analyzing longitudinal clinical scores in such RCT scenarios.

The NLMMRM approach assumes that the trajectory of the response in a specific study can be described by a longitudinal model. However, this methodology doesn't require that the model used (Equation 1) should be the only possible model. Any other model that adequately characterizes the longitudinal scores (HAMD or other) in the trial following the administration of an antidepressant drug can be used with this approach.

The performance of the NLMMRM approach and the comparison of this method with the conventional MMRM analysis was conducted using data from five RCTs which tested paroxetine in a total of 1837 MDD patients from 124 recruitment centers. NLMMRM used the center-specific level of placebo response as a weighting factor in the evaluation of TE. The underlying assumption of this analysis is that centers with high placebo response are less informative than the others for estimating the 'true' treatment effect. As a consequence the weight of the information generated in centers with high placebo response should be lower than the weight of the information generated in the centers with low, or more 'normal,' placebo response.

In this analysis each recruitment center was considered as an independent source of information characterized by its own level of noise defined by the level of placebo response. The overall TE was considered as the resultant of the weighted TE estimated in each recruitment center. Within a given RCT, centers with high placebo response will proportionally contribute less in detecting the clinical efficacy of an antidepressant treatment, than the centers with low placebo response.

A similar situation is described by the signal detection theory, when the signal and the background noise are inversely correlated.

A center's performance was defined as the probability of detecting a signal of a treatment effect. This signal was defined by a clinically relevant separation between active treatment response and placebo response. In this framework, the level of placebo response was considered as a confounding factor (noise) that conditions the probability of signal detection.

The basic assumption in this analysis is that the centers with high placebo responses do not allow to detect an efficacy signal, and that combining these centers with others dilutes the treatment effect, and can sometimes render the overall multicenter trial failed. An important issue associated with this assumption is whether this occurrence of centers with high placebo response is randomly or is nonrandomly distributed. To address this point we have explored the distribution of the efficacy signal (the treatment effect) by center. The evaluation has been done by using the Wald–Wolfowitz test (Wald and Wolfowitz, 1943), also

known as the Runs test for randomness. This test is used to test the hypothesis that a series of numbers (in our case the value of the treatment effect) is random. A run is a set of sequential values that are either all above or below the mean. To simplify computations, the data were first centered about their mean. To carry out the test, the total number of runs was computed along with the number of positive and negative values. A positive run is then a sequence of values greater than zero, and a negative run is a sequence of values less than zero. We can then test if the number of positive and negative runs is equally distributed. The results of the analysis clearly indicated that the TE is not randomly distributed. In addition, we explored the distribution of this value across the different centers in a given study and we found that such a distribution is not normal and is different study by study.

A legitimate question associated with the application of the proposed methodology concerning the potential risk of inflating the Type I error. To discharge this risk, we considered the worst case scenario where the weight associated with the uninformative centers was set to 0. This corresponds to excluding these centers from the analysis and to applying the band-pass filter approach where only the informative centers are used for estimating TE (Merlo-Pich *et al*, 2010). The risk of false positive results associated with the use of this methodology was evaluated by using a clinical trial simulation (Gomeni and Merlo-Pich, 2012). The results of the analysis indicated the band-pass filter methodology preserved the Type I error rate irrespective of whether band pass filtering approach was implemented. Based on these results, we can expect no estimation bias of the proposed methodology by just down weighting uninformative centers, considering no bias has been shown in worst case scenario of excluding such uninformative centers.

One of the major advantages of the NLMMRM method is that it is based on a model. This approach, on one hand, enables to perform clinical trial simulation for evaluating the performances of different study design (including the evaluation of optimal study design) and, on the other hand, allows to account for the level of placebo response in the estimation process. One limitation with this approach is that the NLMMRM method requires a longitudinal model. Therefore, the appropriate model has to be developed and validated prior to implement this approach. However, it has historically been possible to develop such longitudinal models characterizing the disease severity scores.

The main limitation of this study may be the restricted number of RCTs used to evaluate the performances of MMRM and NLMMRM. While the analyses evaluated five RCTs with large number of subjects, the results need to be replicated in other collections of clinical trials with different study designs, inclusion and exclusion criteria, placebo treatment durations, arm numbers, and dates of execution. The authors believe that such an undertaking will only lead to further evidence of the utility of this approach.

In conclusions, the proposed NLMMRM approach provides a critical tool to control the confounding effect of high placebo response, to increase signal detection and to provide a more reliable estimate of the 'true' treatment effect by controlling false negative results associated with excessively high placebo response. This is of particular relevance when decision for investment on progressing into phases IIb–III for a New Chemical Entity is based on the efficacy signal detected in proof-of-concept trials.

FUNDING AND DISCLOSURE

This study has been sponsored by Pharmacometrica France. The authors declare no conflict of interest and that they did not receive any financial support for this study. RG is a full time employee of Pharmacometrica. NG is a full time employee of GSK. He receives salary and stocks from GSK. FB is a consultant of Pharmacometrica. MF: *Research Support:* Abbot Laboratories; Alkermes; American Cyanamid; Aspect Medical Systems; AstraZeneca; Avanir Pharmaceuticals; BioResearch; BrainCells; Bristol-Myers Squibb; CeNeRx BioPharma; Cephalon; Clintara, LLC; Covance; Covidien; Eli Lilly and Company; EnVivo Pharmaceuticals; Euthymics Bioscience; Forest Pharmaceuticals; Ganeden Biotech; GlaxoSmithKline; Harvard Clinical Research Institute; Hoffman-LaRoche; Icon Clinical Research; i3 Innovus/Ingenix; Janssen R&D, LLC; Jed Foundation; Johnson & Johnson Pharmaceutical Research & Development; Lichtwer Pharma GmbH; Lorex Pharmaceuticals; Lundbeck; MedAvante; Methylation Sciences; National Alliance for Research on Schizophrenia & Depression (NARSAD); National Center for Complementary and Alternative Medicine (NCCAM); National Institute of Drug Abuse (NIDA); National Institute of Mental Health (NIMH); Neuralstem, Inc.; Novartis AG; Organon Pharmaceuticals; PamLab, LLC.; Pfizer; Pharmacia-Upjohn; Pharmaceutical Research Associates; Pharmavite LLC; PharmoRx Therapeutics; Photothera; Reckitt Benckiser; Roche Pharmaceuticals; RCT Logic, LLC (formerly Clinical Trials Solutions, LLC); Sanofi-Aventis US LLC; Shire; Solvay Pharmaceuticals; Stanley Medical Research Institute (SMRI); Synthelabo; Wyeth-Ayerst Laboratories. *Advisory/ Consulting:* Abbott Laboratories; Affectis Pharmaceuticals AG; Alkermes; Amarin Pharma; Aspect Medical Systems; AstraZeneca; Auspex Pharmaceuticals; Bayer AG; Best Practice Project Management; BioMarin Pharmaceuticals; Biovail Corporation; BrainCells; Bristol-Myers Squibb; CeNeRx BioPharma; Cephalon; Cerecor; CNS Response; Compellis Pharmaceuticals; Cypress Pharmaceutical; DiagnoSearch Life Sciences (P); Dinippon Sumitomo Pharma; Dov Pharmaceuticals; Edgemont Pharmaceuticals; Eisai; Eli Lilly and Company; EnVivo Pharmaceuticals; ePharmaSolutions; EPIX Pharmaceuticals; Euthymics Bioscience; Fabre-Kramer Pharmaceuticals; Forest Pharmaceuticals; GenOmind, LLC; GlaxoSmithKline; Grunenthal GmbH; i3 Innovus/Ingenis; Janssen Pharmaceutica; Jazz Pharmaceuticals; Johnson & Johnson Pharmaceutical Research & Development, LLC; Knoll Pharmaceuticals; Labopharm; Lorex Pharmaceuticals; Lundbeck; MedAvante; Merck; MSI Methylation Sciences; Naurex; Neuralstem; Neuronetics; NextWave Pharmaceuticals; Novartis AG; Nutrition 21; Orexigen Therapeutics; Organon Pharmaceuticals; Otsuka Pharmaceuticals; Pamlab, LLC.; Pfizer; PharmaStar; Pharmavite LLC.; PharmoRx Therapeutics; Precision Human Biolaboratory; Prexa Pharmaceuticals; Puretech Ventures; PsychoGenics; Psylin Neurosciences; RCT Logic, LLC (formerly Clinical Trials Solutions, LLC); Rexahn Pharmaceuticals; Ridge Diagnostics; Roche; Sanofi-Aventis US LLC.; Sepracor; Servier Laboratories;

Schering-Plough Corporation; Solvay Pharmaceuticals; Somaxon Pharmaceuticals; Somerset Pharmaceuticals; Sunovion Pharmaceuticals; Supernus Pharmaceuticals; Synthelabo; Takeda Pharmaceutical Company Limited; Tal Medical; Tetragenex Pharmaceuticals; TransForm Pharmaceuticals; Transcept Pharmaceuticals; Vanda Pharmaceuticals. *Speaking/Publishing:* Adamed; Advanced Meeting Partners; American Psychiatric Association; American Society of Clinical Psychopharmacology; AstraZeneca; Belvoir Media Group; Boehringer Ingelheim GmbH; Bristol-Myers Squibb; Cephalon; CME Institute/Physicians Postgraduate Press; Eli Lilly and Company; Forest Pharmaceuticals; GlaxoSmithKline; Imedex, LLC; MGH Psychiatry Academy/Primedia; MGH Psychiatry Academy/Reed Elsevier; Novartis AG; Organon Pharmaceuticals; Pfizer; PharmaStar; United BioSource; Wyeth-Ayerst Laboratories. *Equity Holdings:* Compellis; PsyBrain. *Royalty/patent, other income:* Patent for Sequential Parallel Comparison Design (SPCD), which are licensed by MGH to Pharmaceutical Product Development, LLC (PPD); and patent application for a combination of Ketamine plus Scopolamine in Major Depressive Disorder (MDD). Copyright for the MGH Cognitive & Physical Functioning Questionnaire (CPFQ), Sexual Functioning Inventory (SFI), Antidepressant Treatment Response Questionnaire (ATRQ), Discontinuation-Emergent Signs & Symptoms (DESS), and SAFER; Lippincott, Williams & Wilkins; Wolkers Kluwer; World Scientific Publishing.

REFERENCES

Alexopoulos GS, Kanellopoulos D, Murphy C, Gunning-Dixon F, Katz R, Heo M (2007). Placebo response and antidepressant response. *Am J Geriatr Psychiatry* **15**: 149–158.

Allan LG, Siegel SA (2002). A signal detection theory analysis of the placebo effect. *Eval Health Prof* **25**: 410–420.

Beal S, Sheiner LB, Boeckmann A, Bauer RJ (2009). *NONMEM User's Guides. (1989-2009), Icon Development Solutions*. Ellicott City: MD, USAMD, USA.

Benedetti F, Mayberg HS, Wager TD, Stohler CS, Zubieta JK (2005). Neurobiological mechanisms of the placebo effect. *J Neurosci* **25**: 10390–10402.

Chen YF, Yang Y, Hung HM, Wang SJ (2011). Evaluation of performance of some enrichment designs dealing with high placebo response in psychiatric clinical trials. *Contemp Clin Trials* **32**: 592–604.

Fava M, Evins A, Dorer D, Schoenfeld D (2003). The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. *Psychother Psychosom* **72**: 115–127.

Fava M, Mischoulon D, Iosifescu D, Witte J, Pencina M, Flynn M *et al* (2012). A double-blind, placebo-controlled study of aripiprazole adjunctive to antidepressant therapy among depressed outpatients with inadequate response to prior antidepressant therapy (ADAPT-A Study). *Psychother Psychosom* **81**: 87–97.

Gomeni R (2014). Use of predictive models in CNS diseases. *Curr Opin Pharmacol* **14**: 23–29.

Gomeni R, Merlo-Pich E (2007). Bayesian modeling 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.

Gomeni R, Merlo-Pich E (2012). Trial Simulation to estimate Type I error when a population window enrichment strategy is used to improve efficiency of clinical trials in depression. *Eur Neuropsychopharmacol* **22**: 44–52.

Harrell FE (2001). *Regression Modeling Strategies*. Springer-Verlag: New YorkNew York.

Hedlund J, Vieweg B (1979). The Hamilton Rating Scale for Depression. *J Operat Psychiatry* **10**: 149–165.

Heger M (2013). Trial designs advance to overcome bitter pill of placebo effect. *Nat Med* **19**: 1353.

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.

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

Khan A, Schwartz K, Kolts RL, Ridgway D, Lineberry C (2007). Relationship between depression severity entry criteria and antidepressant clinical trial outcomes. *Biol Psychiatry* **62**: 65–71.

Kilts CD, Wade AG, Andersen HF, Schlaepfer TE (2009). Baseline severity of depression predicts antidepressant drug response relative to escitalopram. *Expert Opin Pharmacother* **10**: 927–936.

Kobak KA, Kane JM, Thase ME, Nierenberg AA (2007). Why do clinical trials fail?: the problem of measurement error in clinical trials: time to test new paradigms? *J Clin Psychopharmacol* **27**: 1–5.

Mallinckrodt CH, Lane PW, Schnell D, Peng Y, Mancuso YP (2008). Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. *Drug Inf J* **42**: 303–319.

Merlo-Pich E, Alexander RC, Fava M, Gomeni R (2010). A new population enrichment strategy to improve efficiency of placebo-controlled clinical trial in depression. *Clin Pharmacol Ther* **88**: 634–642.

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.

Nguyen DV, Şentürk D, Carroll RJ (2008). Covariate-adjusted linear mixed effects model with an application to longitudinal data. *J Nonparametr Stat* **20**: 459–481.

Nierenberg AA, Wright EC (1999). Evolution of remission as the new standard in the treatment of depression. *J Clin Psychiatry* **60**: 7–11.

Papakostas GI, Shelton RG, Zajecka JM, Bottiglieri T, Roffman J, Cassiello C *et al* (2014). Effect of adjunctive L-methylfolate 15 mg among inadequate responders to SSRIs in depressed patients who were stratified by biomarker levels and genotype: results from a randomized clinical trial. *J Clin Psychiatry* **75**: 855–863.

SAS Institute Inc *SAS/STAT® User's Guide, Version 9.3*. SAS Institute Inc: Cary, NCCary, NC (2010).

Stein DJ, Baldwin DS, Dolberg OT, Despiegel N, Bandelow B (2006). Which factors predict placebo response in anxiety disorders and major depression? An analysis of placebo-controlled studies of escitalopram. *J Clin Psychiatry* **67**: 1741–1746.

Targum SD, Burch DJ, Asgharnejad M, Petersen T, Gomeni R, Fava M (2014). Use of band-pass filter analysis to evaluate outcomes in an antidepressant trial for treatment resistant patients. *Eur Neuropsychopharmacol* **24**: 1188–1195.

Targum SD, Little JA, Lopez E, DeMartinis N, Rapaport M, Ereshefsky L (2012). Application of external review for subject selection in a schizophrenia trial. *J Clin Psychopharmacol* **32**: 825–826.

Targum SD, Pollack MH, Fava M (2008). Re-defining affective disorders: relevance for drug development. *CNS Neurosci Therapeut* **14**: 2–9.

Trivedi MH, Rush AJ (1994). Does a placebo run-in or a placebo treatment cell affect the efficacy of antidepressant medications? *Neuropsychopharmacology* **11**: 33–43.

Wald A, Wolfowitz J (1943). An exact test for randomness in the non-parametric case based on serial correlation. *Ann Math Stat* **14**: 378–388.

Walsh BT, Seidman SN, Sysko R, Gould M (2002). Placebo response in studies of major depression: variable, substantial and growing. *JAMA* **287**: 1840–1847.