Leveraging magnetic resonance imaging - annualized relapse rate relationship to aid early decision making in multiple sclerosis clinical drug development

Mathangi Gopalakrishnan[1*], Mukul Minocha[1], Joga Gobburu[1]

[1]Center for Translational Medicine, University of Maryland, Baltimore, MD, USA

***Corresponding author:**
Mailing address: Center for Translational Medicine, University of Maryland, 20, N. Pine Street, Baltimore, MD-21201, USA. E-mail: mgopalakrishnan@rx.umaryland.edu

Abstract

Currently, go/no-go decision making in proof-of-concept (POC) multiple sclerosis (MS) trials for promising drug/dose selection are predominantly qualitative in nature. POC trials employ placebo corrected magnetic resonance imaging lesion counts (MRI-T2 counts) as endpoint, whereas, phase 3 trials employ annualized relapse rate at 24 months (ARR-24) as the efficacy endpoint. The objective of the current investigation is to provide a quantitative framework that can aid informed decision making in MS clinical drug development. Blinded summary level data on MRI-T2 lesions at 12 months and aggregate ARR-24 across six clinical development programs digitized from a Food and Drug Administration (FDA)'s 2012 science day presentation were utilized to develop a pharmaco-statistical model linking the MRI-T2 lesions at 12 months with ARR-24. The developed MRI-T2-ARR-24 model was further evaluated by clinical trial simulations and was used to predict the probability of phase 3 clinical trial success given the MRI results in POC trial. The MRI-T2-ARR-24 model suggested that for a unit increase in the MRI-T2 counts, the mean predicted ARR-24 increased by 10%. The model correctly predicted the trial outcomes of four out of the six published MS trials with individual trial predicted ARR-24 values within ± 60% bias. Clinical trial simulations indicated that at least 60% reduction in MRI-T2 counts from placebo in proof-of-concept trials (at any dose or regimen) is needed to achieve a minimum of 80% probability of technical success in the phase 3 trial. Given the competitive landscape in the MS drug development, the decision tool kit could aid in reducing the failure rate in MS phase 3 trials and provide a quantitative framework for more informed dose selection. Further it is anticipated that for significant formulation changes post approval, the MRI-T2-ARR-24 model may be used for bridging efficacy (ARR-24) based only on MRI-T2 data.

Keywords: Multiple sclerosis, Annualized relapse rate, MRI lesions

1.0 Introduction

Multiple Sclerosis (MS) is a disabling immune-mediated disease caused by disruption of the myelin sheath that protects the neurons. Approximately 400,000 people are affected by MS in the United States and 2.5 million worldwide (Noonan, Kathman, & White, 2002). Drug development for relapsing remitting MS (RRMS), as with any other neurological disease, faces numerous challenges with a 35% registration trial failure rate for CNS drugs (Kola & Landis, 2004). Disease modifying trials for RRMS typically involve comparing the novel treatment against a control treatment, i.e., placebo. Currently, proof of concept (POC) and/or dose ranging trials in MS employ biomarkers such as magnetic resonance imaging (MRI) brain lesion activity measured by T1 and T2 weighted images and gadolinium-enhanced lesions (GEL) as trial endpoints (G. Comi *et al.*, 2010). These T1 or T2 lesions depict the demyelination in neurons that is a characteristic of RRMS. Phase 3 trials, however, employ annualized relapse rate at 24 months (ARR-24) as the clinical endpoint. Thus, the decision to proceed to a considerably large efficacy trial for a novel compound (or different doses for the same compound) is based qualitatively on MRI results from POC trials, without the quantitative understanding of what these POC results might translate to in the Phase 3 trials. There exists no quantitative basis to project the probability of the efficacy trial success rates at the selected dose(s) based on MRI data generated early in clinical development. Given the highly competitive nature of this market, there is a grave need for a selection criterion to differentiate lead compounds early in development. Quantitative approaches such as modeling and simulation allow the integration and analysis of biomarker-clinical endpoint data collected across drug development programs (Gobburu & Marroum, 2001). We, therefore, initiated this project to quantify the relationship between early phase MRI-T2 lesions (biomarker) and ARR-24 (clinical endpoint), using the data that were presented at 2012 Food and Drug Administration (FDA) science day (Owen, Jain, Zhang, & Zineh, 2013). In this investigation, we utilize the graphical summary level data from the FDA presentation, from six Phase-3 trials. The objectives were a) to develop a pharmaco-statistical model that correlates MRI-T2 counts at 12 months to ARR-24. b) to evaluate the model by comparing the predicted trial outcome with the published Phase 3 trial results and c) to demonstrate how the developed biomarker-clinical endpoint relationship can facilitate clinical screening of novel compounds and aid as a quantitative decision toolkit for making early phase go/no-go decisions for selected dose(s) or compounds. It is hypothesized that this quantitative decision toolkit may potentially lead to reduced Phase 3 trial failure rates in RRMS, and aid in an optimal screening of compounds based on MRI-T2 results. With respect to regulatory application for this model, matching of T2 lesion counts at 6 or 12 months for a supplemental or extension of approved new drug applications (NDA) (such as an extended release formulation or modified dosing regimen) is proposed as a basis for approval (surrogate marker) in place of ARR-24 as the clinical endpoint.

2.0 Methods

2.1 Data

Data was digitized from the FDA's 2012 Science Day presentation (Figure 1) using Get Data Graph digitizer software (http://getdata-graph-digitizer.com/). The graphical summary level data consisted of over 5500 patients (1150 in placebo and 4381 in active treatment) across six clinical trials that had new or enlarging MRI-T2 lesions counts (ranging from 0-10) and relapses till duration of the study. Aggregate ARR-24 (ranging from 0.19 to 0.7) for the subjects with the same T2 count was provided in the presentation. T2 counts greater than 10 were assigned MRI count as 10. The aggregate ARR-24 (here after will be referred to as ARR-24) was calculated by adding the number of relapses over the treatment duration and normalizing it by 365 days (yearly). A relapse is confirmed by a physician and depicts the disability progression in a particular subject.

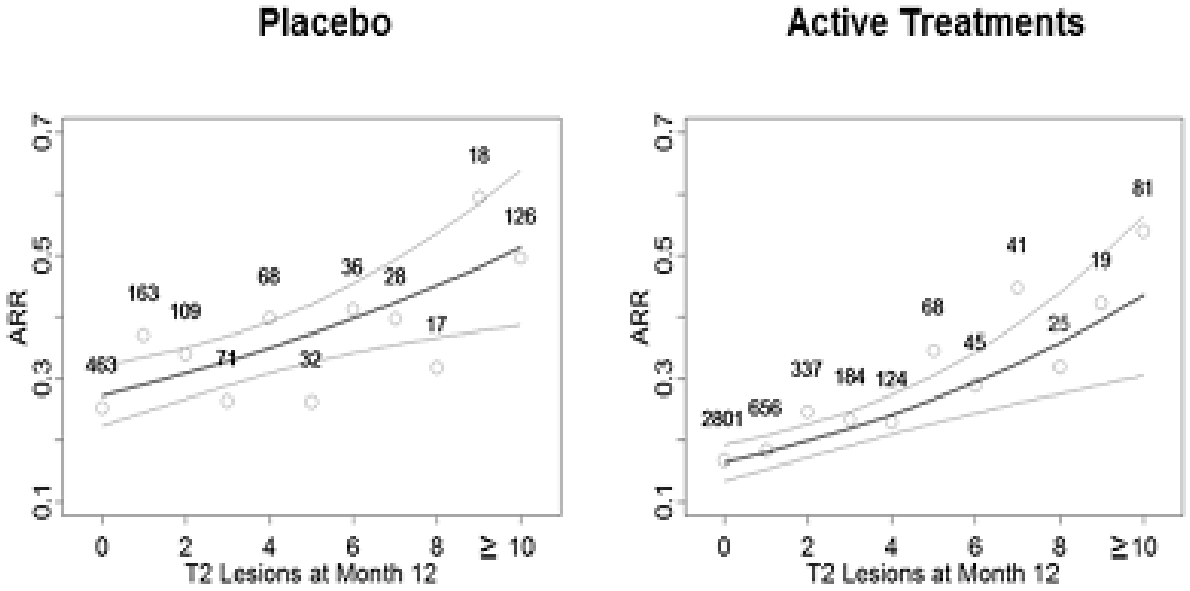

Figure 1: Data obtained from 2012 FDA Science day presentation (Clin Pharmacol Ther. 2013 Jun; 93(6):471-3)

2.2 Development of MRI-T2 (imaging biomarker) - ARR-24 (clinical endpoint) model

A log-linear model was developed to describe the relationship between ARR-24 MRI-T2 at 12 months assuming a negative binomial (NB) distribution for the number of relapses (Y. C. Wang, Meyerson, Tang, & Qian, 2009). The log-linear model that is assumed is shown below:

$$\log(\mu_{ARR-24}) = \beta_0 + \beta_1 \times \text{MRI } T2 \text{ counts} \quad \text{Eq (1)}$$

Where, β_0 is the intercept or the baseline potential to have a relapse and β_1 is the difference in the logarithm of ARR-24 for a unit change in MRI lesion counts, or rate ratio parameter. Since subject level data regarding the number of relapses and the treatment duration were not available, the negative binomial regression was performed directly on ARR-24 with no offset term. The intercept (β_0) and the rate ratio (β_1) parameters of the model were

estimated. The treatment (active treatment or placebo) was not tested as a covariate due to an assumption of treatment independence in this relationship. For the estimated model parameters and predicted model, 95% Wald confidence intervals (CI) were obtained.

2.3 Evaluation of MRI-T2 (imaging biomarker) - ARR-24 (clinical endpoint) model

The developed MRI-T2-ARR-24 model was evaluated through clinical trial simulations by comparing the predicted ARR-24 results (assuming several placebo-corrected MRI-T2 results in POC trials) with the published results for RRMS treatments (Cohen *et al.*, 2010; Fox *et al.*, 2012; Miller *et al.*, 2007; O'Connor *et al.*, 2011; Rudick *et al.*, 2006). Briefly, a compound specific Phase 3 trial was simulated using the mean T2 lesion count data and the associated variability (for both active treatment and placebo) from previously published POC or Phase 2 trials (G. Comi *et al.*, 2008, 2010; Kappos *et al.*, 2008; O'connor *et al.*, 2006). The T2 lesion counts were simulated assuming the NB distribution, and then, each individual's ARR-24 was predicted based on the MRI-T2-ARR-24 model (Eq (1)). The predicted ARR-24 (for placebo and active treatment) and of the placebo-corrected ARR-24 (calculated from subtracting the ARR-24 for placebo subjects from that on active treatment for each compound) for each Phase 3 trial was calculated. This constituted one replicate of the simulation process. The process was repeated 1000 times yielding 1000 estimates of the predicted ARR-24. The mean of these 1000 estimates was then computed and compared to the observed results of ARR-24 in published trials. The uncertainty or 95% CI (for the point estimate of predicted ARR-24) was obtained using 2.5^{th} and the 97.5^{th} percentile of the predicted ARR-24 values from 1000 trial replications. The variability in these simulations were driven by reported variability in MRI-T2 that was assumed to follow NB distribution.

In addition to the comparison at mean level between predicted ARR-24, and observed results, the predicted trial outcome was also compared with the observed trial outcome. The null hypothesis of no difference in the mean of predicted $(\widehat{\mu_{ARR-24}})$ARR-24 between active treatment and the placebo was tested in trial simulations. For this purpose, statistical significance of slope of the treatment effect (H_0:α_1 =0) was assessed for each trial using a NB regression model (shown below) for the predicted ARR-24.

$$\log(\widehat{\mu_{ARR-24}}) = \alpha_0 + \alpha_1 \times TRT \qquad \text{Eq (2)}$$

Consequently, relative bias in the predicted ARR-24 was also calculated using equation below.

$$relative\ bias\ (\%) = \left[\frac{predicted\ ARR24\ -\ observed\ ARR24}{observed\ ARR24}\right].100$$

2.4 Trial simulations to predict the probability of Phase 3 trial technical success

To evaluate the utility of the developed MRI-T2 and ARR-24 relationship to aid in early decision making, clinical trial simulations were performed to predict the probability of Phase 3 trial technical success assuming several phase 2 trial results. Trial simulations consisted of two parts as described below.

2.4.1 Simulation of Phase 2 (MRI) trials

An early phase MRI trial data was simulated with 70 subjects per arm (placebo and active treatment group), which is the typical sample size for a Phase 2 setting. New or enlarging T2 counts were simulated assuming a NB distribution for placebo and treatment arm separately using the mean T2 lesion counts (range as observed in the literature) (Kappos *et al.,* 2008). Similar over-dispersion parameter approximating to a standard deviation (SD) of twice the mean T2 lesion count in placebo and treatment arm was utilized. For the base case scenario, a 50% reduction in T2 lesion counts in the active treatment arm (relative to placebo) was assumed. For example, if the mean T2 lesions were 2.5 in the placebo, a 50% reduction will assume absolute mean T2 lesion counts to be 1.25 in the active treatment arm at the end of a trial. The above simulation procedure was adopted for different scenarios, where the treatment effects were varied ranging from 10% to 75% reduction in mean T2 counts in the treatment group relative to placebo.

2.4.2 Probability of Phase 3 trial success given phase-2 trial results

mean and the associated SD for T2 lesion counts (of the Phase 2 trial simulated above) was now utilized to further simulate the larger Phase 3 trial. The methodology was similar as explained in section 2.3, except the T2 lesion counts were simulated for 400 subjects (instead of specific number in published trials) which is a typical sample size for a Phase 3 trial in treatment and placebo arm separately. Hypothesis testing of no difference in the mean predicted ($\widehat{\mu_{ARR-24}}$) ARR-24between the treatment and placebo group was conducted as described, in section 2.3 (Eq (2)). If the slope of the model fit was statistically significant (i.e., p-value <0.05), the trial was deemed successful, else a failure. Each scenario was repeated 300 times with every replicate sampling a new mean T2 lesion count from the associated uncertainty distribution. The probability mof Phase 3 trial technical success was obtained by calculating the mean of the 300 success or failures of the predicted ARR-24 in the Phase 3 trial. This approach was employed for all the scenarios for the phase-2 trial results ranging from 10% to 75% reduction in mean T2 lesion relative to placebo. An overview of the trial simulation setup is depicted in Figure 2. All the simulations and statistical analysis were performed in R software (R Core Team, 2015). Further, sensitivity analysis was carried out to verify whether assumed placebo mean T2 lesion count influenced the overall trial outcome. The placebo mean T2 lesion counts tested ranged from 1.8 to 3.0 based on the published clinical trial results for the previously FDA-approved drugs for treatment of MS (Cohen et al., 2010; Fox et al., 2012; Miller et al., 2007; O'Connor et al., 2011; Rudick et al., 2006).

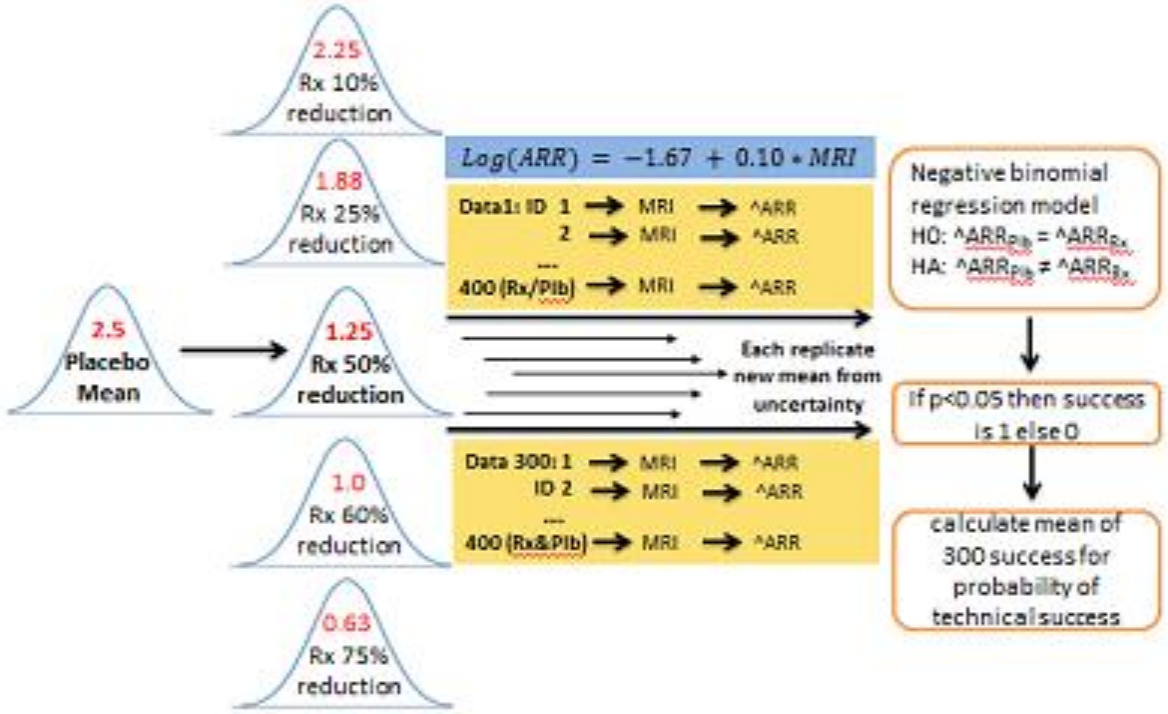

Figure 2 Schematic showing the trial simulations with uncertainty in mean T2 lesion counts at 12 months to predict the probability of Phase 3 trial technical success

3.0 Results

3.1 Data

The graph that was used to digitize the data from the FDA 2012 science day presentation is shown in Figure 1. The black solid line represents the mean predicted ARR-24 at a given T2 lesion counts and the grey lines represent the 95% CI band for the mean predicted ARR-24. These predictions were model based that was developed at the FDA; however, the model parameters were not provided in the presentation. The number on the graph represents the number of subjects with corresponding MRI counts in the placebo and active treatment. Summary statistics of the data (Figure 1) that was digitized from the 2012 FDA science day presentation is shown in Table 1. Data were available from 4381 patients on active treatment and 1150 patients on placebo collected across six Phase 3 trials. The T2 lesion counts were over-dispersed with a mean of 2.65 and 1.02 and associated standard deviation (SD) of 3.4 and 2 in the placebo and active treatment group respectively. The trial and the compound names were undisclosed in the presentation. The aggregate ARR-24 (y-axis) was calculated by averaging the number of relapses for all patients having the same T2 lesion count on placebo and active treatment. For example, an ARR-24 of 0.19 (y-axis, figure 1) represents the grand mean for all relapses for all 2801 subjects with zero T2 counts (x-axis, figure 1) divided by the total duration on treatment, normalized to 365 days. The ARR-24 had a mean of 0.33 and 0.20 and associated SD of 0.09 and 0.07 for the placebo and active treatment group respectively.

Table 1. Summary statistics of the digitized data available from the 2012 FDA science day presentation and compared with the data summarized from the pivotal trials published in the literature

ENDPOINT	FDA DATA; MEAN (SD)		PREVIOUSLY PUBLISHED PHASE 3 DATA; MEAN (SD)	
	PLACEBO	RX	PLACEBO	RX
MRI-T2	2.65 (3.39)	1.02 (2.01)	2.64 (3.33)	0.94 (2.30)
ARR-24	0.33 (0.09)	0.20 (0.07)	0.33 (0.09)	0.20 (0.07)

RX: Active treatment

3.2 MRI-T2 (imaging biomarker) - ARR-24 (clinical endpoint) model

Assuming a NB distribution of the number of relapses over 24 months, the model correlating the MRI-T2 at 12 months and the clinical endpoint, ARR-24 indicates that for a unit increase in the MRI-T2 lesion counts, there is 10% increase in the mean predicted ARR-24.

$$(Eq.1: \widehat{\beta_0}: -1.67\,(95\%\,CI: -1.74, -1.61),\ \widehat{\beta_1}: 0.10\,(95\%\,CI: 0.08, 0.12))$$

Figure 3 shows that the model reasonably describes the observed data. The solid line and the gray area represents the mean model predicted ARR-24 and the associated 95% CI band at a given T2 lesion counts ranging from 0-10.

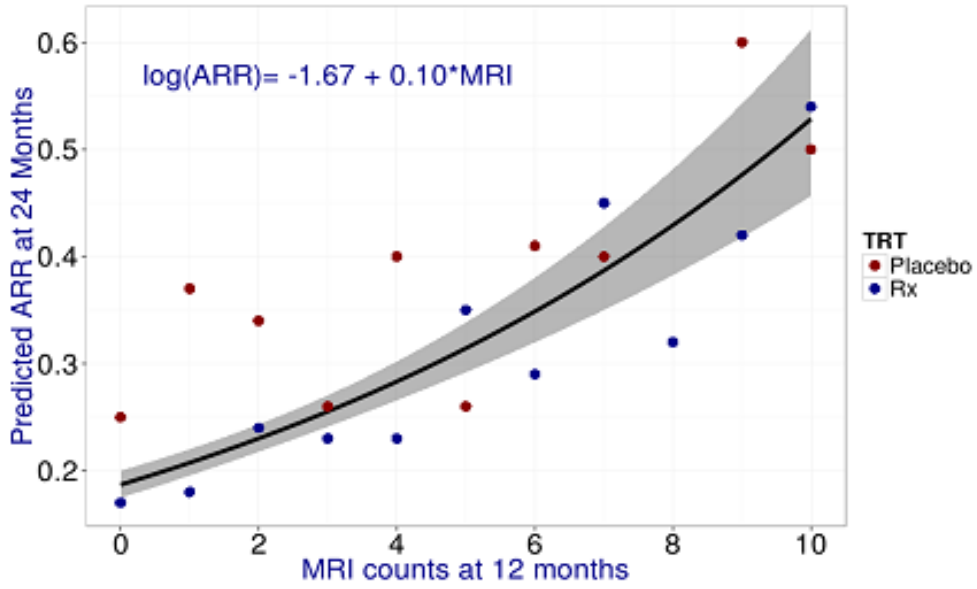

Figure 3. Estimation of MRI-T2 lesion counts and ARR-24 relationship

3.3 Evaluation of MRI-T2 (imaging biomarker)-ARR-24 (clinical endpoint) model

In order to evaluate the MRI-T2-ARR-24 relationship, model predicted mean ARR-24 values for the active and the placebo treatment was compared with the observed Phase 3 ARR-24 results. Table 2 shows that four out of six Phase 3 trial results (67%) were predicted correctly, i.e., statistically significant reduction in ARR-24 on active treatment compared to placebo ($p<0.05$), with individual trial predicted ARR-24 values within ± 60% bias (Table 2). In addition, placebo-corrected ARR-24 values were calculated. Figure 4 shows that the in all six cases, placebo-corrected ARR-24 is under-predicted, which can be considered as a conservative result with regards to future trial simulations and prediction of outcomes. In other words, assuming a smaller magnitude in placebo-corrected ARR-24 may lead to fewer cases of making an incorrect 'go decision' for a molecule based on POC results.

Table 2 Evaluation of the MRI-T2-ARR-24 model by comparison of observed and predicted trial outcomes

Trial	Observed T2 counts Month 6/12		ARR observed		ARR predicted		Relative Bias (%)		P-value[3]
	RX[1]	PLB[2]	RX	PLB	RX	PLB	RX	PLB	
Fingolimod	1.7	2.6	0.18	0.33	0.28	0.35	58	5	<0.05
Natalizumab-1	1.2	6.1	0.23	0.73	0.26	0.59	13	-19	<0.05
Natalizumab-2	0.5	2.4	0.34	0.75	0.20	0.32	-41	-57	<0.05
Dimethyl fumarate	2.2	4.2	0.20	0.40	0.32	0.44	60	11	<0.05
Teriflunomide	0.4	1.5	0.37	0.54	0.20	0.23	-44	-58	0.46
Laquinimod	1.6	2.4	0.30	0.39	0.26	0.30	-11	-23	0.2

[1]RX: Active Treatment, [2]PLB: Placebo, [3] A significant p-value ($p<0.05$) indicates that treatment group was shown superior to placebo based on the mean predicted ARR-24 values obtained from the MRI-T2-ARR-24 model. Two natalizumab Phase 3 trial results were available.

3.4 Trial simulations to predict the probability of Phase 3 trial technical success

Figure 2 depicts the overall process used to predict the probability of Phase 3 trial success given different POC MRI trial results. The probability of Phase 3 trial success was calculated from 300 trial replications for each scenario. Figure 5 shows that the success rate increases as the placebo-corrected MRI effect (percent reduction in T2 lesion counts from placebo) increases from 10% to 75%. Trial simulations suggested that at least 60% reduction in new and enlarging T2 counts compared to placebo is needed to achieve 80% probability of Phase 3 trial success (trial success defined as statistically significant reduction in mean ARR-24 compared to placebo). Sensitivity analysis indicated that the placebo mean for T2 lesion counts below 2.5 significantly reduced the registration trial success probability (Figure 5).

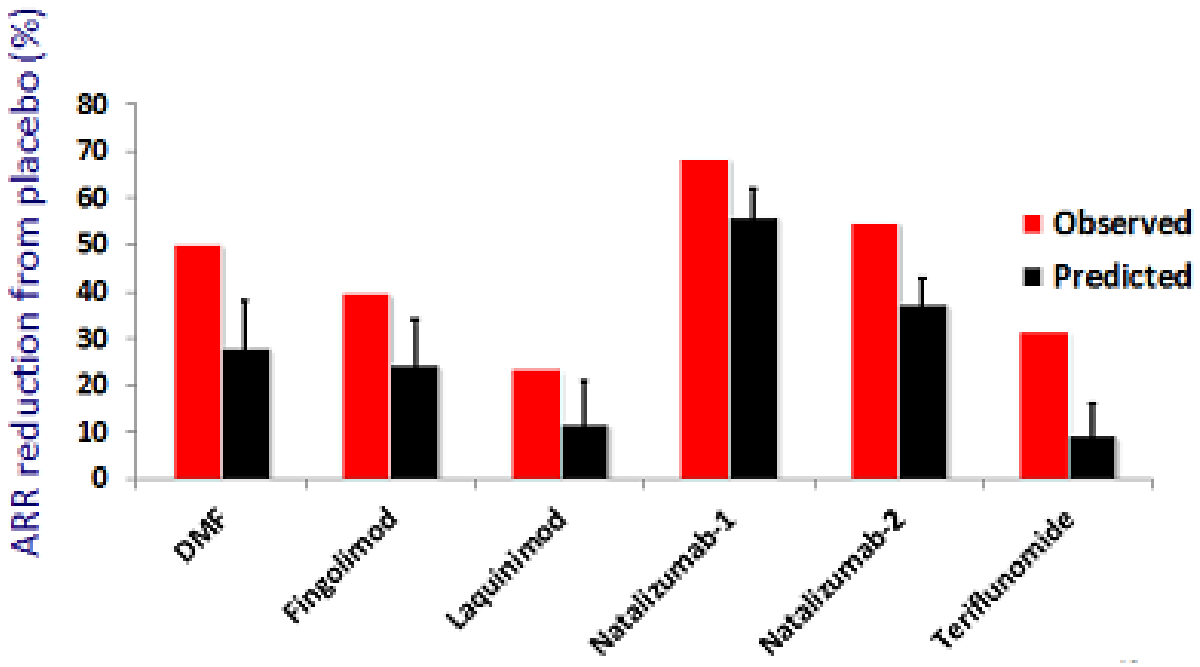

Figure 4. Observed and predicted ARR-24 reduction from placebo in percentage (95% CI) for previously published trials

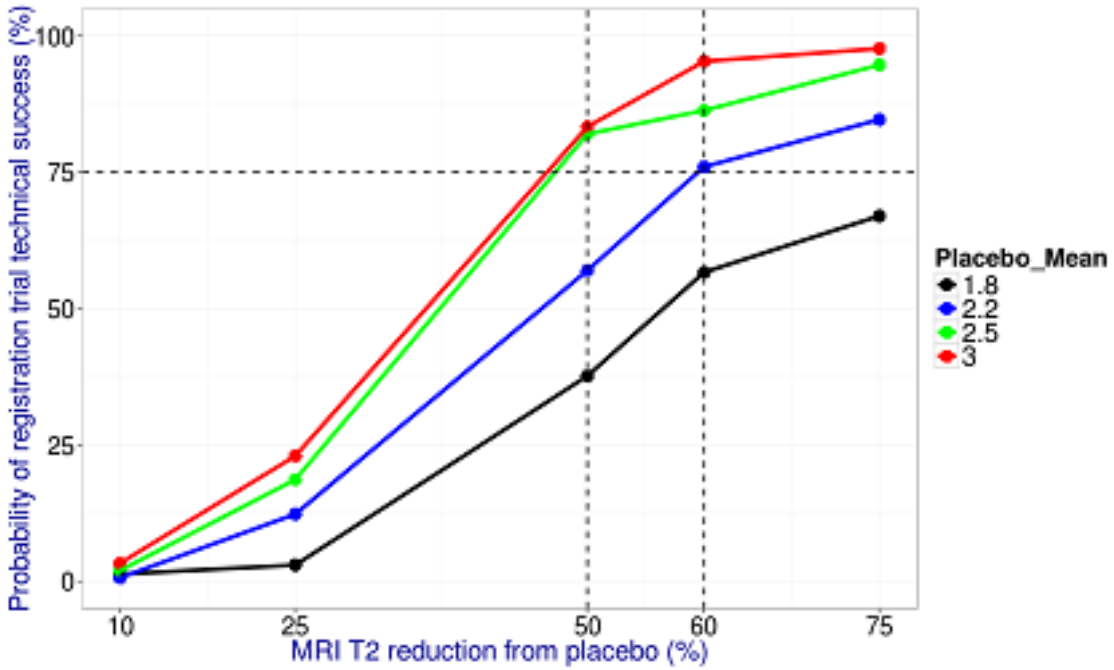

Figure 5. Probability of registration (Phase 3) trial success given different POC placebo corrected MRI-T2 lesions and at varying placebo mean T2 lesion counts.

4.0 Discussion

This investigation is the first attempt to quantitatively characterize the new and enlarging MRI-T2 lesion counts as a predictor of ARR-24. The MRI-T2-ARR-24 model was evaluated by comparing the predicted Phase 3 trial outcomes (placebo corrected ARR-24; magnitude and statistical significance) with the previously published clinical trials. Until now, a majority of stakeholders in MS development landscape use qualitative basis to support dose/regimen selection based on POC trial using MRI-T2 counts as a trial endpoint. We believe that this decision toolkit provides an opportunity to use quantitative and statistical basis to inform key MS clinical drug development decisions such as molecule screening and dose selection. In addition, we try to make a case for the placebo corrected MRI-T2 effect to be used as a potential surrogate of ARR-24 for the approval of extensions for an already approved NDA. We itemized our approach into three objectives. The first objective was to quantify the MRI-T2 (imaging biomarker) - ARR-24 (clinical endpoint) relationship by developing a pharmaco-statistical model. The developed MRI-T2 - ARR-24 model suggests that a unit increase in the T2lesion at 12 months will lead to 10% increase in the mean ARR-24 at 24 months (Figure 3).

Although the visual inspection of the estimated MRI-T2 - ARR-24 model indicated different trajectories for the placebo and treatment group; given an underlying assumption for treatment independence of a

true biomarker-clinical endpoint relationship, treatment was not included as a covariate in this model, similar to the reported tumor size-survival relationship (Y. Wang et al., 2009). The discrepancy (for two apparent relationships for treatment and placebo) can be due to lack of information for individual relapses, time spent on treatment and difference in the data density for the two groups. Nevertheless, given that this data was collected across six phase 3 trials, this is the richest resource of information available thus far. Another key aspect of this model is the presence of an intercept. An intercept suggests that subjects with zero T2 lesion count, will have a minimum predicted ARR-24 of 0.19 (reflective of the observed data in figure 1). This indicates that the T2 count may not be the sole predictor of ARR-24. However, due to lack of information, we can only speculate that some other factors (demographic or baseline covariates) might also be playing a role as a predictor of relapses (and hence ARR-24).

The second objective was to evaluate the developed MRI-T2-ARR-24 model by comparing the predicted ARR-24 and the predicted trial outcome with the published phase 3 trial results. The outcomes of four out of six phase 3 trials were predicted reasonably well with respect to the magnitude and statistically significant reduction in predicted ARR-24 in treatment arm compared to the placebo arm of the respective trials (Table 2). The reason for the discrepancy between the predicted and observed results of two trials (teriflunomide and laquinimod) is explained later in this section. We also computed the percent ARR-24 reduction from placebo for each trial (Figure 4). The published trial data ranged from 23% (for laquinimod) to 68% reduction in placebo corrected ARR-24 (for natalizumab) trials, whereas, the model predicted results ranged from 9% (for teriflunomide) to 56% (for natalizumab). It is evident from figure 4 that the model consistently under-predicted the ARR-24 reduction relative to placebo in comparison to the observed trial results.

We believe that this is a conservative result as using the MRI-T2-ARR-24 model could minimize the chances of going forward with compounds or dose which were actually ineffective in reducing ARR-24 in comparison to placebo. On the other hand, it can be argued that this might also lead to instances of stalling the development of a compound or dose which are moderately effective. Given the competitive nature, we believe it is advantageous to not to develop moderately effective drugs but better or more effective drugs than the comparators. Relative prediction bias was also computed in predicted ARR-24 values in placebo and treatment arms separately (Table 2). All predicted ARR-24 values were within ±60% bias, which in our opinion is acceptable; given T2 counts at 12 months were being used as the sole predictor variable for ARR-24. Overall, these results indicated that the MRI-T2- ARR-24 model was able to predict the previously published trial outcomes reasonably well. The third objective was to construct a go/no-go decision tree or a

decision tool kit by estimating how much percent reduction in T2 counts compared to placebo was 'good enough' to further characterize the compound in subsequent clinical trials. In addition, we wanted to predict the probability of phase 3 trial success assuming different POC trial results (ranging from 10% reduction to 75% reduction in T2 counts relative to placebo at 6 months). Following assumptions were incorporated in the relationship and trial simulations. First, we assumed that the T2 lesion counts at 6 months were similar to T2 counts at 12 months. The relationship was based on T2 counts at 12 months but the typical duration of a POC trial is 6 months. Hence, it is important to make a go/no-go decision based on 6 months data to gain advantage with respect to overall development timeline. The assumption is based on the report that was presented at PAGE meeting in 2011 (Mercier, 2011). The authors reported that another imaging biomarker, GEL does not change drastically after four to five months of active treatment. Hence, it is reasonable to assume that new and enlarging T2 counts also will not change drastically from 6 months to 12 months. Due to the exponential nature of the relationship and the NB distribution of T2 counts, the subjects with simulated T2 counts greater than 10 would attain unrealistic predicted ARR-24 values. Hence, the second assumption was that for those subjects the predicted ARR-24 will be restricted to 2. This assumption was based on the fact that the proportion of patients with simulated counts >10 ranged from 3-5% and so does the proportion of patients with 4 relapses (maximum observed in previous trial) leading to an ARR-24 of 2. Third assumption is that none of the patients would have a predicted ARR-24 of zero, as even a zero T2 lesion count will lead to a minimum predicted ARR of 0.19 due to the presence of intercept in the relationship. Fourth and final assumption was the over dispersion parameter was assumed to be the same for both placebo and treatment arms. The over dispersion parameter value was chosen such a way that would approximate the associated SD to be twice of the assumed mean T2 count in both placebo and treatment arms. Figure 5 depicts the probability of registration trial success given varying placebo corrected MRI-T2 lesion count at six months. The trial simulations suggested that the a 50% reduction in the T2 counts in the POC trial (at 6 months) will achieve approximately 80% probability of Phase 3 trial technical success if the mean MRI T2 lesion counts in the placebo group was 2.5. A placebo mean of 2.5 was selected initially for the base case scenario (50% reduction in mean T2 counts versus placebo) due to the range of placebo mean T2 counts observed in published trials (ranging from 1.5 to 6 and median being 2.4). We also wanted to investigate the influence of mean placebo T2 count on the overall trial outcome by varying the T2 placebo mean in either direction of 2.5. Sensitivity analysis revealed that a placebo mean of 3.0 did not alter the overall trial outcome (Figure 5). However, when a placebo mean of 2.2 was selected, at the same placebo corrected MRI-T2 lesion count of 50%

reduction in comparison to placebo, the probability of technical success dropped from 80% (as predicted with 2.5) to 55%. This value even dropped further to 40% with the placebo mean T2 count being 1.8. This analysis suggested that the value of mean MRI- T2 lesion counts assumed for the placebo group is a sensitive parameter and the results obtained using the relationship needs to be interpreted with caution, when the placebo mean T2 count is below 2.5. The lower probability of technical success resulting from lower mean placebo T2 counts can be further explained as follows: As the placebo mean is assumed to be less than 2.5, the proportion of subjects simulated with zero T2 lesion counts (at six months) would increase substantially and more so in the treatment arm, wherein the mean T2 lesion counts would be even lesser as compared to placebo mean. In addition, due to the presence of intercept, increased proportion of zero T2 counts (simulated) will lead to individual predicted ARR-24 values to be 0.19 in both placebo and treatment arm, thus leading to statistically insignificant result upon comparison of predicted ARR-24 between the two arms. We believe this phenomenon also explains why teriflunomide trial outcomes was not predicted to be significant even with a 75% reduction T2 lesion count, as the placebo and treatment mean T2 counts were 1.5 and 0.4 respectively (Table 2) . For laquinimod, since the percent reduction in T2 lesion count itself was less than 33%, and the magnitude of placebo corrected ARR-24 was being under-predicted, simulations did not predict the overall trial outcome to be significant.

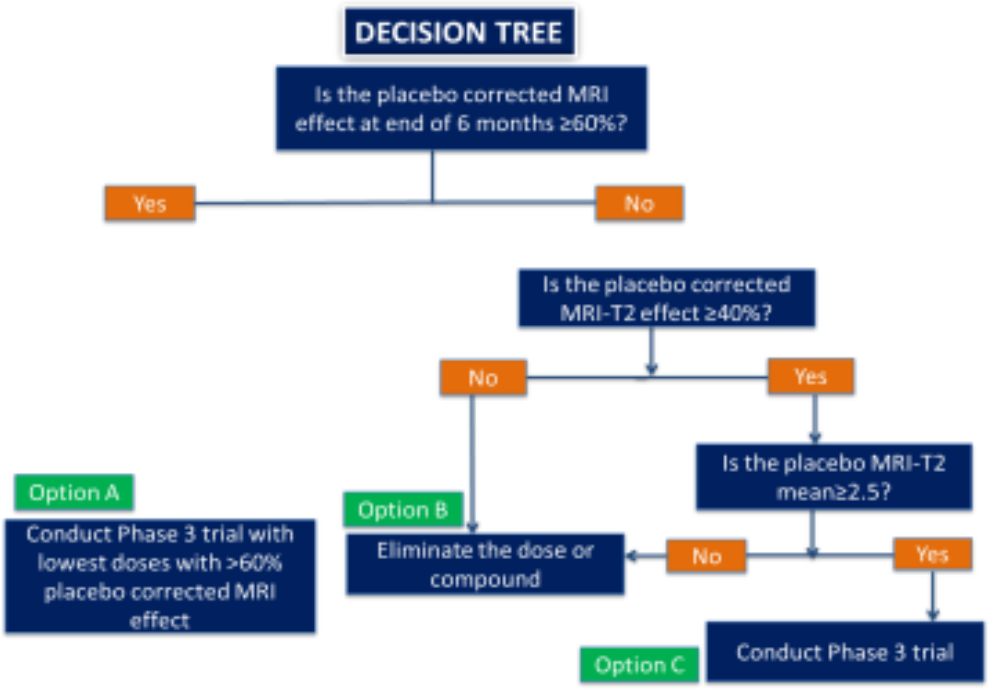

Figure 6. Go/no-go decision tree to aid compound screening and dose selection in MS drug development.

With regards to the go/no-go decision tree, the first question needs to be asked is whether the placebo corrected MRI effect for the investigative drug in the POC trial is greater than or equal to 60%. If the answer is 'yes' then the company may invest in the molecule (option A, Figure 6). However, if the answer is 'no', then a second question that needs to be asked is whether the placebo corrected MRI effect is at-least 40%? If the answer is 'no' then there is no advantage in further investing in the compound (option B) as the competitor drugs are already available with such an effect. However, if the answer is 'yes' then given the moderate effect, only if a major safety or adherence benefit over previously approved drugs exists, the compound could be further characterized (option C). Figure 7 represent the observed placebo corrected MRI effect and the corresponding observed ARR-24 effect relative to placebo from published trials.

For natalizumab and teriflunomide, the placebo corrected MRI-T2 effect is greater than 60% (first question in decision tree) and for dimethyl fumarate (DMF) and fingolimod, although the placebo corrected MRI-effect was moderate (50%), but there was an advantage of oral administration compared to injectable drug natalizumab. Although the registration trial has been conducted for laquinimod, it is worth noting that this drug has not been approved and had an observed ARR-24 reduction from placebo of just 25% given a 33% reduction in T2 lesion counts at 6 months (Giancarlo Comi et al., 2012). This visual check demonstrates that given the assumptions and the limitation of under-prediction of ARR-24 in our research, this decision toolkit can still be used as a viable guiding option for making informed decisions at critical stage of MS clinical drug development.

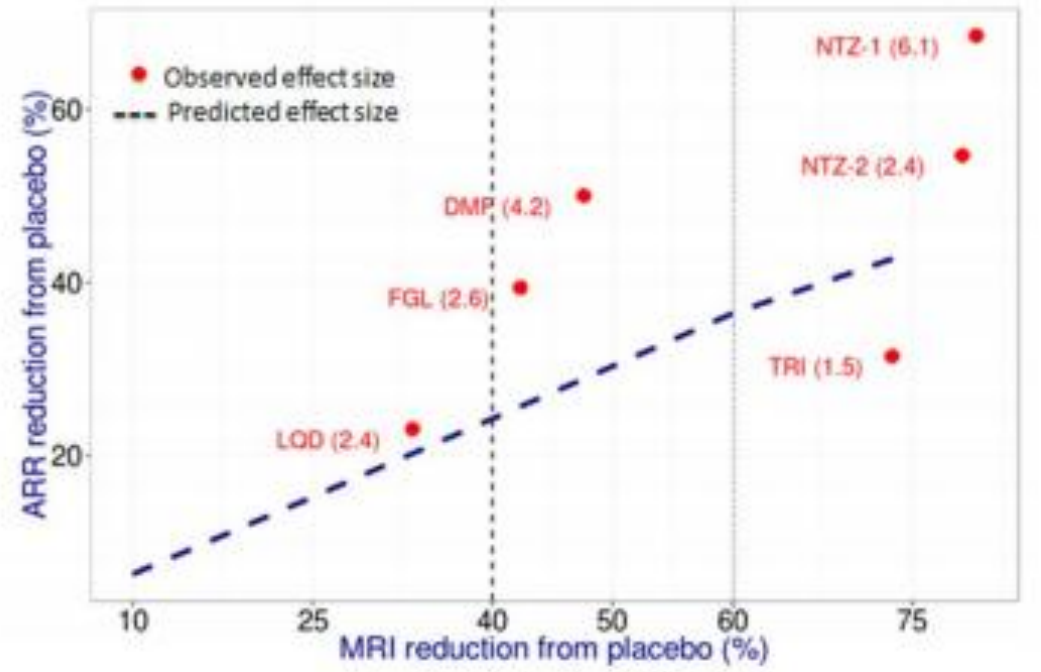

Figure 7. Previously conducted MRI and corresponding Phase 3 trial results. The red symbols are the observed data with reported T2 lesion counts in parenthesis and the blue dotted line represents the predicted ARR-24 reduction from placebo given varying phase-2 trial results (based on MRIT2-RR-24 model). LQD:Laquinimod, FGL: Fingolimod, DMF: Dimethyl fumarate, NTZ: Natalizumab, TRI: Teriflunomide

Finally, along with this decision tree (Figure 6) we also wanted to make a case for using placebo corrected MRI effect as a surrogate for clinical endpoint (ARR-24) for extensions of the approved NDAs. For example, if the new dosing regimen could match the placebo corrected MRI effect for the previously approved dose at six months, is it really necessary to demonstrate reduction in ARR-24 in a 24 month trial? To answer this question we compared the success rate for both POC trial and Phase 3 trial with respective study designs (70 and 400 subjects per arm for MRI and ARR-24 trials) at the same placebo corrected MRI effect (ranging from 10-75% reduction in T2 lesions relative to placebo). Results indicate that both the curves predict similar respective success rates at any given placebo corrected MRI effect, (Figure 8) meaning, MRI trial success can be a viable predictor of the ARR trial success as well. Given the drug is the same and only the dosing regimen has changed, the magnitude of the ARR-24 effect should not be different for the previously approved dose provided the placebo corrected MRI effect has been matched. This approach can also be extended to modified release formulations and even biosimilars. Currently, European Medical Agency has already started employing MRI endpoints (T1 and T2 lesion counts) as the basis of approval of biosimilars for previously approved interferons (European Medical Agency, 2011). Apart from biosimilars, this toolkit can also open up an avenue for approval of already approved drugs in adults in pediatric population, (currently 2-5% of total MS patients worldwide) although more research needs to be focused on finding if the disease progression is similar between adults and pediatrics.

In conclusion, we acknowledge there are several assumptions and limitations associated with the MRI-T2 -ARR-24 model and trial simulations methodology, however, given the richness of information collected from the registration trials at the FDA, this is currently the best option available to provide quantitative rationale to support dose selection in early MS drug development.

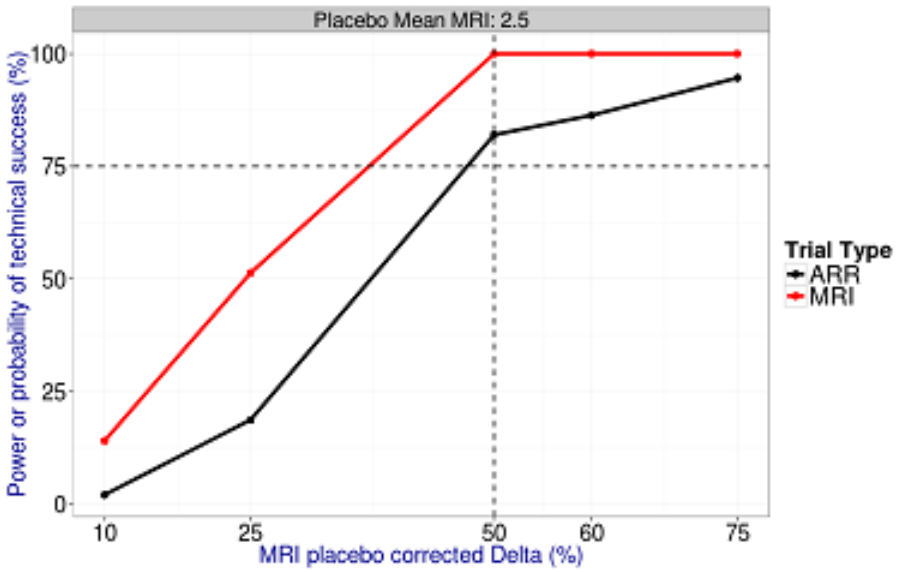

Figure 8. MRI trial success is a good predictor of ARR trial success. Red curve represents the power to detect a significant reduction in T2 lesion counts whereas the black curve represents the probability of phase 3 trial technical success given same reduction in T2 counts relative to placebo (%).

5.0 References

1. Cohen, J. A., Barkhof, F., Comi, G., Hartung, H.-P., Khatri, B. O., Montalban, X.,TRANSFORMS Study Group. (2010). Oral fingolimod or intramuscular interferon for relapsing multiple sclerosis. *The New England Journal of Medicine, 362(5),* 402 415.http://doi.org/10.1056/NEJM oa0907839

2. Comi, G., Jeffery, D., Kappos, L., Montalban, X., Boyko, A., Rocca, M. A., ALLEGRO Study Group. (2012). Placebo-controlled trial of oral laquinimod for multiple sclerosis. *The New England Journal of Medicine,* 366(11), 1000-1009. http://doi.org/10. 1056/NEJMoa1104318

3. Comi, G., O'Connor, P., Montalban, X., Antel, J., Radue, E.-W., Karlsson, G., ..FTY720D2201 Study Group. (2010). Phase II study of oral fingolimod (FTY720) in multiple sclerosis: 3-year results. *Multiple Sclerosis,*16(2),197-207.http://doi.org/ 10.1177/1352458509357065

4. Comi, G., Pulizzi, A., Rovaris, M., Abramsky, O., Arbizu, T., Boiko, A., ..others. (2008). Effect of laquinimod on MRI-monitored disease activity in patients with relapsing- remitting multiple sclerosis: a multicentre, randomised, double-blind, placebo-controlled phase IIb study. *The Lancet,* 371(9630), 2085-2092.

5. European Medical Agency, 2011. (2011). Draft guideline on similar biological medicinal products-WC 500120652. pdf. Retrieved September 1, 2015, from http:// www.ema.europa. eu/docs/en_GB/document_library/Scientific_guideline/2012/01/WC500120652.pdf

6. Fox, R. J., Miller, D. H., Phillips, J. T., Hutchinson, M., Havrdova, E., Kita, M., CONFIRM Study Investigators. (2012). Placebo-controlled phase 3 study of oral BG-12 or glatiramer in multiple sclerosis. The New England Journal of Medicine, 367(12), 1087-1097.http://doi.org/10.1056/NEJMoal 206328

7. Gobburu, J. V., & Marroum, P. J. (2001). Utilisation of pharmacokinetic pharmacodynamic modelling and simulation in regulatory decision-making. *Clinical Pharmacokinetics,* 40(12), 883-892.http://doi.org/10.2165 /00003088-200140120 00001

8. Kappos, L., Gold, R., Miller, D. H., MacManus, D. G., Havrdova, E., Limmroth, V., others.(2008). Efficacy and safety of oral fumarate in patients with relapsing-remitting multiple sclerosis: a multicentre, randomised, double-blind, placebo-controlled phase IIb study. *The Lancet,* 372(9648), 1463-1472.

9. Kola, I., & Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? *Nature Reviews. Drug Discovery,*3(8),711-715.http://doi.org/ 10.1038/nrd1470

10. Mercier, F. et al. (2011). *A Bayesian meta-analysis of longitudinal lesion count in multiple sclerosis patients.* PAGE, Athens, Greece. Retrieved from http://www.pagemeeting.org/pdf

assets/1382FMercier%20PAGE2011%20Poster.pdf

11. Miller, D. H., Soon, D., Fernando, K. T., MacManus, D. G., Barker, G. J., Yousry, T. A., AFFIRM Investigators. (2007). MRI outcomes in a placebo-controlled trial of natalizumab in relapsing MS. *Neurology,* 68(17), 1390-1401.http://doi.org/10. 1212/01 .wnl.0000260064.77700.fd

12. Noonan, C. W., Kathman, S. J., & White, M. C. (2002). Prevalence estimates for MS in the United States and evidence of an increasing trend for women. *Neurology,* 58(1), 136-138.

13. O'connor, P. W., Li, D., Freedman, M. S., Bar-Or, A., Rice, G. P. A., Confavreux, C., others. (2006). A Phase II study of the safety and efficacy of teriflunomide in multiple sclerosis with relapses. *Neurology,* 66(6), 894-900.

14. O'Connor, P., Wolinsky, J. S., Con favreux, C., Comi, G., Kappos, L., Olsson, T. P., TEMSO Trial Group. (2011). Randomized trial of oral teriflunomide for relapsing multiple sclerosis. *The New England Journal of Medicine,365(14),*12931303.http://doi. org/10.1056/NEJMoa101 4656

15. Owen, R. P., Jain, L., Zhang, L., & Zineh, I. (2013). Office of clinical pharmacology science day: a forum to stimulate innovation in clinical pharmacology. *Clinical Pharmacology and Therapeutics,* 93(6), 471-473. http://doi.org/10.1038/clpt.2013.33

16. R Core Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (Version 3.1). Retrieved from http://www.R-project.org/.

17. Rudick, R. A., Stuart, W. H., Cala-bresi, P. A., Confavreux, C., Galetta, S. L., Radue, E.-W., SENTINEL Investigators. (2006). Natalizumab plus interferon beta-1a for relapsing multiple sclerosis. *The New England Journal of Medicine,* 354(9), 911-923. http://doi.org/10.1056/NEJMoa044396

18. Wang, Y. C., Meyerson, L., Tang, Y. Q., & Qian, N. (2009). Statistical methods for the analysis of relapse data in MS clinical trials. *Journal of the Neurological Sciences, 285(1-2),*206-211.http://doi.org/10.101 6/jjns.2009.07.017

19. Wang, Y., Sung, C., Dartois, C., Ramchandani, R., Booth, B. P., Rock, E., & Gobburu, J. (2009). Elucidation of Relationship Between Tumor Size and Survival in Non-Small-Cell Lung Cancer Patients Can Aid Early Decision Making in Clinical Drug Development. *Clinical Pharmacology & Therapeutics, 86(2),* 167-174. http://doi.org/10.1038/clpt.2009.64

