Published in final edited form as:

Mult Scler. 2017 June ; 23(7): 982–987. doi:10.1177/1352458516670733.

Baseline EDSS proportions in MS clinical trials affects the overall outcome and power: a cautionary note

Guoqiao Wang, PhD¹, Gary R. Cutter, PhD², Stacey S. Cofield, PhD², Fred Lublin, MD³, Jerry S. Wolinsky, MD⁴, Tarah Gustafson³, Stephen Krieger, MD³, and Amber Salter, PhD¹

¹Division of Biostatistics, Washington University in St. Louis, St. Louis, USA

²Department of Biostatistics, University of Alabama at Birmingham, Birmingham, USA

³Icahn School of Medicine at Mount Sinai, New York, USA

⁴McGovern Medical School, University of Texas Health Science Center at Houston, Houston, USA

Abstract

Background—In randomized clinical trials, when treatments do not work equally effectively across stratifications of participants, observed event rates may differ from those hypothesized leading to deviations in estimated power.

Corresponding Author: Guoqiao Wang, PhD, Assistant Professor, Division of Biostatistics Washington University in St. Louis, 660 S. Euclid Ave., St. Louis, MO 63110 USA, Phone: (314) 362-0349, Fax: (314) 362-2693, guoqiao@wustl.edu.

Author Contributions: Gary R. Cutter conceived of the idea and designed the study. Guoqiao Wang and Amber R. Salter conducted the analyses and drafted the manuscript. Gary R. Cutter, Stacey S Cofield, Fred Lublin, Jerry Wolinsky, Tarah Gustafson, and Stephen Krieger edited the manuscript. All authors approved of the version to be published.

Disclosures: Dr. Wang: The NARCOMS Research Fellowship is supported in part by the Foundation of the CMSC and an unrestricted educational grant by Biogen.

Dr. Cutter: Consulting and/or DSMB commitments in Past 12 months. Participation of Data and Safety Monitoring Committees: All of the below organizations are focused on medical research: Apotek, Biogen-Idec, Cleveland Clinic(Vivus), Glaxo Smith Klein Pharmaceuticals, Gilead Pharmaceuticals, Horizon Pharmaceuticals, Modigenetech/Prolor, Merck/Ono Pharmaceuticals, Merck, Merck/Pfizer, Neuren, Sanofi-Aventis, Teva, NHLBI (Protocol Review Committee), NINDS, NICHD (OPRU oversight committee). Consulting, Speaking fees & Advisory Boards: Consortium of MS Centers (grant), D3 (Drug Discovery and Development), Genzyme, Genentech, JanssenPharmaceuticals, Klein-Buendel Incorporated, Medimmune, Novartis, Opexa Therapeutics, Receptos, Roche, EMD Serono, Spiniflex Pharmaceuticals, Somahlution, Teva pharmaceuticals, Transparency Life Sciences. Dr. Cutter is employed by the University of Alabama at Birmingham and President of Pythagoras, Inc. a private consulting company located in Birmingham AL. Dr. Cofield: Consulting fees American Shoulder and Elbow Society, DSMB service Medimmune, Grant Review Department of Defense.

Dr. Lubin: Scientific Advisory Board: Acorda Therapeutics Inc.; Actelion Pharmaceuticals Ltd.; Bayer; Biogen Idec; EMD Serono, Inc. (Merck & Co., Inc.); F. Hoffman-La Roche Ltd.; Facilitate, LTD; Forward Pharma; MedImmune, LLC; Novartis; Osmotica Pharmaceutical Corp.; Questcor Pharmaceuticals, Inc.; Receptos Inc.; Revalerio Corporation; Teva Pharmaceutical Industries Ltd.; XenoPort, Inc.; sanofi-aventis; to-BBB. Other Activities: Examples include, but are not limited to, committee participation, data safety monitoring board (DSMB) membership: Gerson Lehrman Group.

Dr. Wolinsky: Consulting fees (AbbVie, Actelion, Alkermes, Athersys, EMD Serono, Forward Pharma, Genentech, Genzyme, Novartis, Roche, Teva, Teva Neuroscience, to-BBB, XenoPort); contracted research (Genzyme, National Institutes of Health, National MS Society through the University of Texas Health Science Center at Houston, Sanofi); royalties (University of Texas Health Science Center at Houston for monoclonal antibodies out-licensed to Chemicon International).

Ms. Gustafson has nothing to disclose.

Dr. Krieger: Consulting/Advisory Board: Acorda Therapeutics; Bayer Healthcare; Biogen Idec; EMD Serono; Genentech, Inc.; Genzyme Corporation; Novartis; Takeda Pharmaceuticals; Teva Pharmaceuticals. Industry-Sponsored Non-Promotional Lectures: Genzyme Corporation; Biogen Idec.

Dr. Salter has nothing to disclose.

Objectives—To investigate the effect of distributions of baseline EDSS proportions in relapsing remitting MS on the trial outcome, confirmed disability progression rate (CDPR), and power.

Methods—We reported CDPRs in the CombiRx trial by baseline EDSS and by groups (1st (0, 1), 2nd (1.5, 2), 3rd (2.5, 3), and 4th (3.5)); and investigated the effect of different combinations of baseline EDSS proportions on the trial outcome and power.

Results—There were 244 (25.4%) participants in the 1st group, 368 (38.4%) in the 2nd, 223 (23.3%) in the 3rd, and 124 (12.9%) in the 4th with CDPRs of 40.1%, 13.9%, 11.2%, and 16.9%, respectively. Both CDPR and power increased when the proportion of the 1st group increased in hypothetical trials with equal sample sizes in each arm, and a 10% increase in the 1st group led to a 5% increase in power.

Conclusions—Various baseline EDSS proportions yielded different CDPRs and power, suggesting caution in interpretation of treatment effects across trials that enrolled participants with different proportions of baseline EDSS.

Keywords

multiple sclerosis; EDSS; RRMS; confirmed disability progression rate

Introduction

The randomized, double-blinded, placebo-controlled design is widely accepted and used in Phase III MS clinical trials.¹ The typical fixed trial design is straightforward to plan and trials of this design can be accurately compared when the treatment works equally effectively across stratifications of participants (e.g. younger vs old, short disease duration vs long disease duration) or when the primary outcome occurs uniformly across the participants.² For example, if the treatment leads to the same changes in the Expanded Disability Status Scale (EDSS) regardless of disease duration, then the trial design is not concerned about the potential recruitment imbalance in terms of disease duration. On the contrary, if disease duration matters, then at the planning stage, the proportion of participants with different disease durations should be taken into account to accurately estimate the power. Once a trial has begun, the proportions may need to be re-estimated to ensure they are in accordance with what was planned. Otherwise, the trial may be underpowered and a type II error could occur, or the trial may be overpowered and a certain amount of resources and time would be wasted. However, when these key assumptions do not hold in a clinical trial, the estimates used for the trial outcome and the associated power may be off target, and consequently lead to the trial being underpowered and the trial failing to detect benefits.

In this paper, we used the CombiRx data to investigate confirmed disability progression (CDP), an increasingly utilized outcome for MS clinical trials, stratified by baseline EDSS level. We demonstrate how violations of the key assumption of the same treatment effect for RRMS participants within the eligible EDSS range affect the trial outcome and power by varying the baseline EDSS proportions in the trial population.

Methods

Data source

The CombiRx trial was a Phase III double-blind, randomized, three-armed, multi-center treatment trial with 1008 RRMS participants comparing 50% of participants on combination therapy (interferon beta 1a and glatiramer acetate) versus 25% on each of the single agents with matching placebo. The duration of the primary study was 36 months. The study cohort included participants 18-60 years of age with a baseline EDSS score ≤ 5.5 . Details of the specific participant inclusion and exclusion criteria, description of the study design, baseline characteristics of the participants, and primary results have been reported previously.^{3, 4} The CombiRx trial was approved by the applicable central or institutional review boards and the data and safety monitoring committee (DSMC) appointed by NIH/NINDS before site initiation and recruitment of participants. Informed consent was obtained prior to any screening procedures or enrollment.

Primary outcome

We defined confirmed disability progression (CDP) as an increase in EDSS score over the baseline EDSS which was sustained for 6 months or 3 consecutive quarterly visits in CombiRx.⁵ The increase in EDSS needed for CDP confirmation varied by baseline EDSS score, where a 1.5 point increase in EDSS from baseline was needed when the baseline EDSS was 0, a 1 point increase for baseline EDSS scores between 1 and 5, inclusively, and a 0.5 point increase when the baseline EDSS was greater than or equal to 5.5. The confirmed disability progression rate (CDPR) was defined as the number of participants with at least one CDP out of the total participants at risk, that is, the percent of participants with CDP at any time over the 3 years. The CDPR was used as the primary outcome for this investigation.

Statistical analysis

First, we described the CDPR in CombiRx by baseline EDSS. For simplicity of our arguments, we divided the study cohort into 4 groups by baseline EDSS scores: 1st (0, 1), 2nd (1.5, 2), 3rd (2.5, 3), and 4th (3.5, 5.5) and calculated the CDPR for each group. To determine whether the 4 groups differed, we used Pearson's chi-square test for categorical variables (CDPRs and sex), and ANOVA for continuous variables (age and total follow-up time). Logistic regression was used to test for differences in the CDPRs of the 4 groups while adjusting for disease duration and age at baseline. All analyses were conducted with SAS 9.4 (SAS Institute Inc., Cary, NC).

Trial Outcome and Power

Since all participants in the CombiRx trial were on treatment with small clinical differences in outcome event rates, we combined participants in the 3 treatment arms to be considered as a single treatment arm. To look at the effect of different proportions of baseline EDSS levels in a trial, we used four combinations of group proportions (Table 1).

For each combination, we calculated the CDPR for the treatment arm. We assumed that an effective treatment would lead to a constant 30% reduction in CDPR over the placebo arm

and the distribution of baseline EDSS scores was equivalent in both arms in a trial with a balanced design (equal sample sizes in each arm), and then we calculated the CDPR of the placebo arm. Next, the absolute reduction in CDPR rate from the placebo arm was calculated for each combination.

Assuming a null hypothesis that the two CDPRs were equal, a type I error of 0.05 for a two-sided test, and a normal distribution for the difference in proportions, we then calculated the power for each hypothetical two-arm trial with 1:1 randomization ratio.

Results

Of the 1008 participants in the CombiRx trial, 959 (95.1%) stayed in the trial for at least 9 months and were eligible for calculating the presence or absence of CDP, the earliest potential time point being 3 months with 6 month confirmation and thus were used for this study. The mean total follow-up time and proportion of females were not significantly different between the four groups, however the mean age was ($p < 0.0001$) (Table 2). As the EDSS level increased, the mean age of each group increased and all pairwise group comparisons were significantly different (all $p \leq 0.001$) (Table 2). The CDPRs were higher in those participants with a baseline EDSS of 0 and 1 compared to the CDPRs for those with higher baseline EDSS levels (Table 2). The overall CDPR over the three year period was 20.3%. Of the 959 participants, 244 (25.4%) belonged to the 1st group, 368 (38.4%) to the 2nd, 223 (23.3%) to the 3rd, and 124 (12.9%) to the 4th, with CDPRs of 40.1%, 13.9%, 11.2%, and 16.9%, respectively. The CDPRs were significantly different among the 4 groups ($p < 0.0001$) and remained so ($p < 0.0001$) after adjusting for disease duration ($p = 0.74$) and age at baseline ($p = 0.0185$). For those in the 1st group, the odds (95% confidence interval) of CDP was 4.5 (3.0-6.7) times those in the 2nd group, 6.0 (3.6-10.0) times those in the 3rd, and 4.0 (2.3-7.0) times those in the 4th group. The odds of CDP were not significantly different among the 2nd, 3rd, and 4th groups.

Applying the CDPRs found in each group from the CombiRx trial to each combination of group proportions, the overall trial CDP increased as the proportion of participants from the 1st group increased (Table 1). Assuming a constant 30% treatment effect, the absolute reduction in CDPR between the treatment and placebo arms also increased with increasing proportions of participants from the 1st group (Table 1). A similar increase in power was observed between the different combinations of group proportions (Figure 1). Given a sample size of 100 in each arm, trials with a combination of 40%, 30%, 20%, and 10% of participants had 43% higher CDPR and 48% more power than those with a combination of 10%, 20%, 30%, and 40%. From a different perspective, for a two-sided test with type I error 0.05, type II error 0.2, and randomization ratio 1:1 to detect a 30% reduction in CDPR, a clinical trial with a combination of 40%, 30%, 20%, and 10% would require approximately 301 participants per arm; while the combination of 10%, 20%, 30%, and 40% would need about 493 per arm. While the differences in CDPR and power increased as participants from the 1st group increased, they remained approximately proportional over the increase in sample size (Table 1 and Figure 1). If the 1st group was excluded such that only groups 2nd, 3rd and 4th were used, the CDPRs among these three more similar groups still varied and

ranged between 11.2% of the 3rd group and 16.9% of the 4th group, a difference of 50% $((16.9-11.2)/11.2)$.

Discussion

In this study, we demonstrated that the CDPR varied by baseline EDSS level, indicating that the outcome of interest may occur unequally across stratifications of participants.

Additionally, we showed that changing baseline EDSS proportions in the trial population would have yielded different CDPRs and thus affected the outcome and power (and correspondingly sample size) of a clinical trial. These facts provide a possible explanation for the differences in the primary outcomes observed across MS clinical trials with similar therapies despite the same or comparable inclusion and exclusion criteria. Different baseline EDSS level proportions may occur under a number of circumstances. For example, trials conducted in different regions may recruit disproportionately from various EDSS groups, especially if one region has a larger proportion of more disabled participants than the other. Small and shorter trials may recruit more from the prevalence pool, being pre-identified during the startup and sustaining the disease for a relatively longer time with higher baseline EDSSs; while larger and longer trials may recruit more from the incidence pool, meaning more participants are newly diagnosed and being eligible for the trial with lower baseline EDSSs. These potential discrepancies in study cohorts are difficult to avoid by only implementing inclusion or exclusion criteria. These variations make both the reproducibility of results, estimates of outcome events, and the sample size estimation for future trials based on previous trials, challenging.

The variation in CDPR by baseline EDSS should also be considered when interpreting or generalizing clinical trial outcomes. Treatment effects demonstrated in clinical trials that enrolled larger proportions of participants with baseline EDSS levels 0 or 1 may have very different results from studies with fewer participants at these levels even though the entrance criteria were identical.

The differences in CDPR by baseline EDSS are also likely to affect sample size re-estimation at interim analyses for modern adaptive designs. An interim analysis where all or most of the participants enrolled had higher baseline EDSS (e.g. more disabled or from the prevalence pool) would then underestimate the CDPR and call for an increase in sample size. If the remaining originally planned number of participants enrolled had lower EDSS scores, then the added samples would not be needed as the CDPR would be closer to what was expected.

More frequently with the lack of head to head trials, across-trial comparisons are made using network analyses and indirectly adjusted comparative analyses. While these will continue to be done out of interest and necessity, our results suggest that the composition of baseline populations may have important impacts and even when adjustments are made on the mean values, the composition of participants even with the same mean can give rise to more or less CDPR and should be another cautionary note in the interpretation of aggregate across-trial comparisons.

There are some limitations of this study. Only one RRMS clinical trial was used to provide estimates of CDPRs, with patients having relatively short time from onset and there is a lack of information available from published RRMS study reports to cross-validate the CDPRs by baseline EDSS. However, the conclusions possess more general implications. As long as the primary outcome may be affected by the baseline characteristics (baseline EDSS in our study) or the treatment does not work equally effectively across stratifications of participant population, our conclusions apply. Additionally, there may be some uncertainty in the generalizability of these results to the progressive MS population. Still, the conclusions of this study apply and may be even more critical for planning primary progressive MS trials given that these trials can only recruit within a smaller EDSS range, typically 3.0 to 6.5.⁶ Additionally, Harding et al. demonstrated that recruitment of participants with only higher baseline EDSS may doom the trial to fail from the beginning,⁶ further emphasizing that in order to avoid failed trials due to design flaws it is important to know how the primary outcome may vary by baseline EDSS levels and then balance the enrollment by baseline EDSS levels accordingly.

Therefore, in planning clinical trials, performing comparisons across clinical trials or estimating and re-estimating sample size, adjustments should be considered not only for demographic characteristics but also for the distribution of the characteristics of interest in the population and not simply whether the mean levels are equal. With CDP as a primary, co-primary or a secondary endpoint becoming more prevalent⁷, reporting of this endpoint by baseline EDSS level is needed. We examined several completed RRMS⁸⁻¹² clinical trials and tried to compare their CDPRs to those in CombiRx by baseline EDSS. However, CDPRs by baseline EDSS were not reported in any of those trials although the overall CDPRs in both the treatment and the control arms were available.⁸⁻¹² This lack of reporting CDPRs by baseline EDSS indicates potential underuse of this important information for planning trials, comparing effectiveness of various treatments, and conducting meta-analyses in RRMS. Additionally, the availability of CDPRs by baseline EDSS coupled with conducting interim analyses provides a potential way to verify the assumptions (i.e. planned power and treatment effects) for future trials. An interim analysis could provide estimates of the proportion of baseline EDSS levels recruited, then using the CDPRs from published study reports, the power and the treatment effect of the ongoing trial can be re-estimated and compared to the pre-planned ones in a blind fashion. If any of the two is deemed inaccurate, then adaptation can be taken to remedy the trial. Overall, our study advocates the necessity and urgency of providing CDPRs by baseline EDSS in future clinical study reports and for consideration of baseline EDSS in the planning of future trials in MS.

Acknowledgments

Funding: NIH/NINDS 5-U01-NS045719. Biogen Idec and Teva Neuroscience generously contributed study medications and matched placebo without input into the design, conduct, analysis, or interpretation of results of the trial.

References

1. Reingold SC, McFarland HF, Petkau AJ. The growing need for alternative clinical trial designs for multiple sclerosis. *Multiple Sclerosis Therapeutics*. 2011; 253

2. Riddell CA, Zhao Y, Petkau J. An adaptive clinical trials procedure for a sensitive subgroup examined in the multiple sclerosis context. *Statistical methods in medical research*. 2013 0962280213480576.
3. Lublin FD, Cofield SS, Cutter GR, et al. Randomized study combining interferon and glatiramer acetate in multiple sclerosis. *Annals of neurology*. 2013; 73:327–40. [PubMed: 23424159]
4. Lindsey J, Scott T, Lynch S, et al. The CombiRx trial of combined therapy with interferon and glatiramer acetate in relapsing remitting MS: design and baseline characteristics. *Multiple sclerosis and related disorders*. 2012; 1:81–6.
5. Cofield SS, Cutter GR, Lublin F, Wolinsky JS, Conwit R. Variability in the EDSS: What does it mean for progression? American Academy of Neurology (AAN), Chicago, IL, USA. 2008:12–9.
6. Harding KE, Wardle M, Moore P, et al. Modelling the natural history of primary progressive multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*. 2015; 86:13–9.
7. Goldman MD, Motl RW, Rudick RA. Possible clinical outcome measures for clinical trials in patients with multiple sclerosis. *Therapeutic advances in neurological disorders*. 2010
8. Cohen JA, Coles AJ, Arnold DL, et al. Alemtuzumab versus interferon beta 1a as first-line treatment for patients with relapsing-remitting multiple sclerosis: a randomised controlled phase 3 trial. *The Lancet*. 2012; 380:1819–28.
9. Comi G, Jeffery D, Kappos L, et al. Placebo-controlled trial of oral laquinimod for multiple sclerosis. *New England Journal of Medicine*. 2012; 366:1000–9. [PubMed: 22417253]
10. O'Connor P, Wolinsky JS, Confavreux C, et al. Randomized trial of oral teriflunomide for relapsing multiple sclerosis. *New England Journal of Medicine*. 2011; 365:1293–303. [PubMed: 21991951]
11. O'Connor P, Filippi M, Arnason B, et al. 250 µg or 500 µg interferon beta-1b versus 20 mg glatiramer acetate in relapsing-remitting multiple sclerosis: a prospective, randomised, multicentre study. *The Lancet Neurology*. 2009; 8:889–97. [PubMed: 19729344]
12. Kappos L, Freedman MS, Polman CH, et al. Long-term effect of early treatment with interferon beta-1b after a first clinical event suggestive of multiple sclerosis: 5-year active treatment extension of the phase 3 BENEFIT trial. *The Lancet Neurology*. 2009; 8:987–97. [PubMed: 19748319]

Figure 1. Power for different combinations of group proportion

Table 1
Combinations of group proportions, CDPRs of treatment groups, and the reduction from placebo groups given a 30% treatment effect

Combination	Group (%)				Overall CDPR of treatment groups	% reduction in CDPR
	1 st	2 nd	3 rd	4 th		
1	10	20	30	40	0.169	7.3
2	20	20	30	30	0.192	8.2
3	30	30	20	20	0.218	9.4
4	40	30	20	10	0.242	10.4

Table 2
Demographics by groups and confirmed disability progression (CDP) in 36 months by baseline EDSS (N total=959)

Groups	Baseline EDSS	Participants by baseline EDSS, N (%)	CDP, N (%)	Age at baseline, mean (SD) *	Total follow-up months, mean (SD) **	Gender (female), N (%) **
1 st	0	119 (12.4)	45 (37.8)	34.5 (8.4)	46.8 (18.5)	177 (72.5)
	1	125 (13.0)	53 (42.4)			
2 nd	1.5	135 (14.1)	26 (19.3)	37.3 (9.0)	44.0 (17.4)	273 (74.2)
	2	233 (24.3)	25 (10.7)			
3 rd	2.5	133 (13.9)	13 (9.8)	39.8 (10.1)	43.0 (17.3)	157 (70.4)
	3	90 (9.4)	12 (13.3)			
4 th	3.5	67 (7.0)	10 (14.9)	43.1 (9.6)	42.0 (18.5)	90 (72.6)
	4+	57 (5.3)	11 (19.3)			

* p-value is <0.0001

** p-values are not significant based on ANOVA or Chi-square test.