Published in final edited form as:

Ann N Y Acad Sci. 2018 January ; 1412(1): 5–13. doi:10.1111/nyas.13501.

Learning from the past: reflections on recently completed MG trials

Michael Benatar¹, James F. Howard Jr.², Richard Barohn³, Gil I. Wolfe⁴, and Gary Cutter⁵

¹University of Miami Miller School of Medicine, Miami, Florida

²University of North Carolina, Chapel Hill, North Carolina

³University of Kansas Medical Center, Kansas City, Kansas

⁴University at Buffalo, State University of New York, Buffalo, New York

⁵University of Alabama Birmingham, Birmingham, Alabama

Abstract

Recently completed clinical trials have varied widely in design, but also perhaps in less explicit ways. We explore ways in which these design characteristics may have influenced recruitment and results, as well as the implications for forthcoming studies. Trial eligibility criteria may inadvertently select for incident versus prevalent cases or patients with relatively mild versus more severe disease. Trial enrichment with patients who have relatively mild disease may limit the sensitivity of the trial to detect a therapeutic effect. Enrichment for patients with more severe disease may introduce confounds caused by regression toward the mean. Overly narrow eligibility may limit the generalizability of results. An exclusive focus on incident cases may hamper recruitment, as may many factors, such as access to the experimental therapeutic treatment outside of the trial or following completion of the double-blind treatment period. We illustrate how other design characteristics (e.g., treatment duration, strategy for steroid tapering, selection of the primary outcome and principal analytic approach) may affect the sensitivity of a trial to demonstrate therapeutic effects. Finally, we consider the importance of placebo effects, being careful to differentiate these from therapeutic effects observed in the placebo group, and discuss how the use of combined outcome measures may minimize placebo effects.

Keywords

clinical trial; eligibility criteria; study design; recruitment; placebo effects

Introduction

The results of four randomized controlled trials (RCTs) of therapeutics for myasthenia gravis (MG) have been published in the 5 years following the 12th International Conference

Address for correspondence: Michael Benatar, 1120 NW 14th Street, CRB1318, Miami, FL 33136. mbenatar@med.miami.edu.

Competing interests

The authors declare no competing interests.

on Myasthenia Gravis and Related Disorders. These include the EPITOME study, a trial of prednisone for ocular myasthenia;^{1,2} the MGTX trial of thymectomy for non-thymomatous MG;³ a trial of methotrexate in patients with generalized MG;⁴ and a phase II trial of eculizumab in patients with treatment-refractory generalized disease.⁵ All but the EPITOME trial, which focused on patients with ocular myasthenia irrespective of antibody status, targeted patients with anti-acetylcholine receptor (AChR)⁺, non-thymomatous, generalized MG. These three trials aimed to address three very different therapeutic niches among the heterogeneous myasthenic population. Three of these trials—EPITOME, MGTX, and eculizumab—yielded positive results, or at least results that warranted further investigation in larger phase III trials. This encouraging success was accomplished despite the fact that all three of these trials struggled to meet recruitment goals. Each of these RCTs contributed meaningfully to the body of evidence supporting the use of various therapies for patients with MG. Reflecting on the designs of these trials, the challenges they faced, and the results they yielded offers an opportunity to enhance future therapeutic development efforts.

Study population and sensitivity to demonstrate therapeutic effect

Trial eligibility criteria typically aim to identify a relatively homogeneous population in which it will be possible to demonstrate a therapeutic effect of the investigational agent. A homogenous population presumably allows for a more consistent response among patients within a treatment group, enhancing the power of the trial by reducing the variability among participants. In the effort to clearly define and achieve this homogeneity within the study population, however, there is potential to restrict eligibility in unintentional ways—either defining an overly narrow segment of the patient population with limited generalizability of results; enriching the study population for patients with relatively mild disease, who are least in need of novel therapies; or selecting a very severe segment of the population and then have responses confounded by regression toward the mean.

Incident versus prevalent cases

Eligibility criteria (summarized in Table 1) aimed to enroll different populations of myasthenic patients. The EPITOME study targeted patients with ocular myasthenia (OM) who had not been treated with prednisone previously. The combination of these two eligibility criteria (OM and steroid naive) essentially limited enrollment to incident cases, because OM patients are often started on steroids by a community neurologist or neuro-ophthalmologist before referral to a neuromuscular specialist. While the MGTX trial aimed to enroll early (< 3 years disease duration initially, which was modified to < 5 years 2 years into the trial to enhance enrollment) patients with anti-AChR⁺, non-thymomatous generalized MG and permitted prior corticosteroid use, eligibility criteria excluded patients who had been treated with other immunosuppressive agents. A consequence (intended or not) was that the MGTX study population was similarly enriched for incident cases and was skewed toward a younger age group (~ 32 years) with relatively short disease duration (~ 12 months). By contrast, average disease duration in the eculizumab trial population was ~ 7 years; this was largely a reflection of the need to enroll a treatment-refractory population, which required failure on previous therapies (i.e., patients with a history of having been treated with at least two immunosuppressive agents at adequate doses for at least 1 year).

While such entry criteria may enhance the ability to detect differences if they exist, they reduce the size of the available population from which to recruit and limit the generalizability of the results. The criteria are nevertheless important to the question. For example, in the thymectomy trial, it was important to know the treatment effect in earlier disease, whereas, for drugs like eculizumab, the need for refractory populations also addresses a different but relevant scientific question related to the treatment.

Disease severity

While eligibility criteria for the methotrexate trial did not specify a quantitative myasthenia gravis (QMG) threshold for enrollment, both the MGTX and eculizumab trials required a QMG score of at least 12, and the eculizumab trial also required a score of 2 on at least four of the QMG items. The net effect was that both the MGTX and methotrexate trials enrolled patients with relatively mild disease (average baseline QMG scores of ~ 12 and 10.5, respectively), while eculizumab enrolled a population with more severe disease (average baseline QMG score: ~ 18) (Table 2). Enrollment of patients with mild disease (reflected by lower QMG scores) raises the possibility of a treatment floor effect (i.e., that there is relatively little room (both in terms of disease severity and range of QMG scores) within which to demonstrate a therapeutic effect). Of note, baseline QMG scores were lowest in the methotrexate trial, which failed to show a clinically and statistically significant benefit of methotrexate. While baseline QMG scores in the MGTX trial were only marginally higher (average ~ 12), a clinically meaningful and statistically significant reduction in QMG scores, albeit modest, was apparent. Interestingly, therapeutic effect as evidenced by a reduction in QMG scores was largest in the eculizumab trial, in which the patient population was characterized by the highest average baseline QMG score. This is not to say that disease severity at enrollment is the sole determinant of sensitivity to demonstrate therapeutic effect, but rather to highlight the potential impact that disease severity and baseline QMG may have on the sensitivity of a trial to show a therapeutic effect. It must also be recognized that lower values of the QMG may not only result in floor effects, but also that discrete point increments or decrements on an ordinal scale will have much greater impact on a low QMG than on a higher QMG. While percent reductions may allow comparisons—since a 2-point change on a QMG starting at 10 is a 20% change, whereas starting at 18, it would take a change of almost 4 points to achieve a similar percent change—MG trials have typically considered absolute reductions in QMG score to define clinically meaningful changes or movement to below a threshold to define “disease control.” Hence, if a trial protocol defines a 3-point change as clinically meaningful, then a 20% change for someone starting at 10 would not achieve the important outcome, whereas a 20% change in someone starting at 18 would. Furthermore, since the QMG scale is ordinal, discerning the difference between scores of 0 (no symptoms) and 1 (mild symptoms) may be more difficult than differentiating between scores of 2 (moderate) and 3 (severe).

Generalizability of results

The trial population in the EPITOME study was small, relatively old with early disease, and not previously exposed to prednisone or other immunosuppressive agents. The generalizability of these results, for example to younger patients with longer disease duration, is unclear. The MGTX study population was younger, but resembled the

EPITOME population insofar as these were largely patients with recent diagnoses who had not yet been treated with immunosuppressive therapy other than corticosteroids. The relevance of the MGTX results, therefore, to an older population with longer disease duration and more prevalent use of non-steroid immunosuppressive therapy or long-term use of steroids is unclear. While the methotrexate trial failed to show a benefit of 12-month treatment in older patients with relatively mild disease, relevance to a more severely affected population with greater steroid requirements is unclear. Similarly, since the eculizumab trial enrolled patients with treatment-refractory disease, relevance to newly diagnosed patients who are relatively treatment naive has yet to be determined.

Recruitment challenges

It has been said that the first rule of clinical trials is recruitment, recruitment, recruitment. All but one of the trials completed within the last 5 years failed to meet recruitment goals (Table 3). While not all sites were open for enrollment for the full duration of each trial, estimates of the average number of patients enrolled per site ranges from a low of 0.81 (eculizumab) to a high of 2.63 (methotrexate); the average number of patients enrolled per month across all activated sites ranges from 0.32 (EPITOME) to 1.75 (MGTX). One possible reason for slow recruitment in the EPITOME and MGTX trials is that eligibility criteria in these trials limited enrollment to incident cases and, in the MGTX study, the willingness of patients to undergo an extended transsternal thymectomy at a time of increasing use of minimally invasive surgery certainly hampered recruitment. Since recruitment cannot begin until a site is activated, and the latter requires contract execution and institutional review board (IRB) approval (among other steps), administrative delays in contracting may have a deleterious effect on recruitment, as was encountered in the EPITOME trial. Multinational trials, such as MGTX, may be especially vulnerable to such delays given the added complexity of international regulatory requirements.⁶ In the MGTX trial, the average number of months to initiate recruitment was almost 12. Such delays may dampen study site and personnel enthusiasm that was present when the site was selected for participation, and this in turn may hamper recruitment. Increasing efforts to utilize central IRB (cIRB) review within the last decade may have slightly lessened delays from duplicative reviews, but multinational studies are limited in their ability to use a cIRB, and these processes still require quite a long lead time.

The methotrexate trial aimed to enroll patients who might benefit from the steroid-sparing effect of methotrexate, but trial eligibility criteria related to prednisone required simply that patients be taking at least 10 mg/day (i.e., a relatively low dose). Median prednisone dose among the trial population was 20 mg/day. The methotrexate trial, therefore, essentially targeted people with relatively mild disease requiring relatively little prednisone. The point is that it may have been challenging to demonstrate a steroid-sparing effect in patients who were taking a relatively low dose of prednisone at the time of enrollment. This is another example of a potential floor effect. Lack of motivation for such patients to enroll in a placebo-controlled trial might have also contributed to slow recruitment. In essence, trial enrollment criteria likely limited the size of the eligible population. Availability of study treatment outside the trial (e.g., prednisone, methotrexate, thymectomy), limited access to study drug(s) not available outside the trial (e.g., eculizumab) after conclusion of the double-

blind trial, and the broader context of a disease for which many effective therapies exist may all have contributed to slow recruitment. Largely absent from the MG literature, however, is a survey of patients to better understand their preferences and motivations for deciding whether to participate in a clinical trial. There may be an opportunity to ask this question of those participating in existing MG patient registries.

It should go without saying that all entry criteria reduce the pool of eligible patients. The lack of carefully crafted estimates of the impact of these criteria on likely enrollment severely affects and delays achieving recruitment goals. The recent push for more pragmatic trials, which aim at generalizability and have few if any exclusions other than for safety reasons, is an attempt to enhance recruitment and make the results more clinically translatable. However, the regulatory issues often conspire to add back in various criteria working against these goals and make the pragmatic trials reflect the real world a bit less.

Impact of study design on sensitivity to demonstrate therapeutic effects

Treatment duration

Duration of treatment and follow-up was 16 weeks in the EPITOME and eculizumab trials, 12 months in the trial of methotrexate, and 3 years in the MGTX trial. Short treatment duration is appropriate in studies, such as the EPITOME and eculizumab trials, that aim to demonstrate short-term symptomatic benefit. Similarly, follow-up in the MGTX trial was appropriately prolonged, as the timing and durability of therapeutic effect of thymectomy were unknown, and it was necessary to provide sufficient time for the dose of concomitant prednisone to be tapered. Follow-up in the methotrexate trial may have been too short given what is known about the steroid-sparing effect of similar agents—azathioprine's effect is observed after 15 months,⁷ and steroid-sparing effects of mycophenolate mofetil were not observed despite a 9-month follow-up period.⁸

The choice of the duration of the study must also weigh costs, feasibility, and the appropriateness to the question. For MGTX, not only was the onset of the therapeutic effect an issue, but also the durability of the treatment effect. For treatments that are not reversible or that will entail long-term exposure where safety and efficacy are important trade-offs, follow-up duration is an important consideration. Where longer-term information is deemed to be extremely valuable, follow-up might need to extend for more than 2 years, pragmatically increasing the overall time needed to complete the study. Often, open label-extension studies are tacked on to enable collection of additional data, but this is most valuable for safety outcomes, although, when a placebo is used during the trial period, reproducing the treatment effects at the crossover can provide data free of placebo and regression to the mean effects if designed appropriately. An alternative approach to the open-label extension is to follow all patients under the protocol until the last patient completes the trial, as was done in the MGTX trial. This enabled MGTX to have follow-up data extending 7 years. However, with only 57 individuals at 5 years of follow-up, the sample size necessary to answer questions about the durability of the responses over the first 3 years may not be adequate.

Steroid-tapering strategy

Both the methotrexate and MGTX trials incorporated protocol-driven steroid-tapering schedules in order to permit detection of steroid-sparing effects. A critical difference between the two studies, however, is that steroid tapering in MGTX required that participants achieve and maintain minimal manifestation status (MMS), whereas the strategy in the methotrexate study only required that participants show clinical improvement on the MGFA postintervention status scale. Targeting greater therapeutic effect (i.e., MMS) is likely to have two consequences: (1) to drive down the QMG score and (2) to require a higher prednisone dose. Indeed, this is precisely what was observed in MGTX compared with methotrexate. Final QMG scores at last follow-up were ~ 6 points in MGTX and ~ 10 points in the methotrexate trial. Similarly, the time-weighted average prednisone dosage was 27 mg/day over 36 months in the MGTX placebo group and 15 mg/day in the methotrexate placebo group. The more stringent requirement of MMS (as opposed to only clinical improvement), therefore, may have kept the doses higher owing to a more rigorous and slow withdrawal protocol and may have made it more likely to empower detection of a therapeutic effect (if indeed it existed) in the MGTX than the methotrexate trial. Irrespective of whether the difference in steroid-tapering strategy between these two trials was the reason for the failure to observe a therapeutic effect in the methotrexate trial, this important difference in study design and drug treatment protocol are significant and relevant to future studies of steroid-sparing therapeutic strategies.

Outcome measures

Primary outcome measures in the methotrexate and eculizumab trials were similar insofar as each relied upon a single outcome—area under the prednisone dose-time curve for methotrexate, and the change in QMG score in the eculizumab trial. By contrast, both the EPITOME and MGTX trials relied on compound outcome measures. The EPITOME study utilized the rate of treatment failure (defined as failure to achieve MMS, development of generalized disease, or discontinuation of study drug because of intolerable side effects). The MGTX trial relied on both symptomatic improvement and steroid-sparing effect and used chose the compound end point to ensure that the clinical outcome achieved was not the result of differential prednisone use. When using clinical outcome measures as the target goal, a single outcome may be more appropriate, but when the amount of drug being used over a defined period is the outcome, it is necessary to ensure that the clinical outcomes are improved or at least the same.

While we are not passing any judgment on the appropriateness of single versus compound outcome measures, this conceptual difference is worth noting and considering explicitly in the design of future trials. Below, we consider the potential impact of outcome measure selection on the magnitude of potential placebo effects.

Statistical methods

The design of trials, and specifically registration trials, generally requires the a priori choice of a primary end point. Regulators want to make sure that a trial is addressing a specific question with defined outcomes to avoid data dredging and wishful thinking. The requirement for a trial to be registered on [ClinicalTrials.gov](https://clinicaltrials.gov) with a priori-specified end

points entails publicly declaring what the organizers of the trials set out to achieve. Thus, a key decision for any trial is choosing the right outcome measure. While too much emphasis may be placed on the resulting P -value from a trial with sentiments like “the trial failed because the P -value was not < 0.05 ” expressed, it is also a reality that choosing the wrong end point can weaken the results of a trial that is possibly positive, but on secondary outcome variables only. In such a situation, it is the coherence and consistency across all the outcomes that should define the success or failure of the trial. The problem or dilemma that has led to the more rigid paradigm of regarding success on the primary as paramount is that it is usually the investigators arguing that the trial is positive only after seeing the data discrepancy. It is not that the investigators are knowingly incorrect, but the evolution of this standard has occurred because it is difficult to distinguish those results that reflect belief bias, where investigators have data dredged to obtain an outcome that is significant, versus results that are true consistent findings of effects irrespective of missing on the P -value and where a different choice on the primary outcome would have yielded evidence of treatment effect. Such may have been the case in the methotrexate study if the QMG had been chosen as the primary end point with prednisone sparing as a secondary end point. Nevertheless, phase II and especially phase III trials require a priori specification of the specific hypotheses and end points. Statisticians often recommend that sample size be based on the end point that requires the largest sample size to show significance, but define the primary outcome as the variable that requires the smallest sample size. This would provide adequate sample size for both primary and secondary end points, thus minimizing the risk of getting stuck with a disconnect between outcomes. Of course, this is not always financially or logistically feasible.

Another statistical concern is the intention to treat paradigm (as randomized) versus per protocol (as treated) approach to analyses. As randomized is relatively clear, but what constitutes per protocol is often defined uniquely in each trial and, all too often, after the trial is over. Again, the focus on this paradigm has been to protect against biased conclusions. But at what cost? Statistically, we are estimating two different outcomes: the intention-to-treat approach estimates effectiveness and the per protocol is aimed at estimating efficacy (but with the likely confounding of informative dropouts). “Per protocol” may also have different definitions. Intention to treat seeks to determine whether people do better if they try to take this treatment, whereas per-protocol analyses address the question of whether people do better if they actually take this treatment. Of course, many things go into a treatment: their ability to tolerate it, how well it works, and many other factors. It is really expected that an intention-to-treat analysis is supported by a per-protocol analysis, so that it is not really a choice of one or the other. In the MGTX trial, all of the results have been reported using the intention-to-treat approach. In the MGTX trial, there were an equal number of crossovers: eight had thymectomy outside of the protocol and eight refused the surgery. The effects of these crossovers has made the reported results conservative. In the extreme, one can see that intention-to-treat analyses alone can be flawed, despite being the accepted and preferred analysis. In the MGTX trial, for example, we could assume that all who were randomized to surgery refused and all who were randomized to prednisone alone demanded surgery. Here, we would find a statistically significant difference, but because of the way the patients were treated, it would be in the opposite direction. Of course, no such

extreme result occurs in practice, but arguing not to perform the intention-to-treat analysis because of this limitation is balanced by the obvious problem with per-protocol analyses. Suppose a disease is fatal to all but a few people who are cured. The per-protocol analysis will declare that this is a successful treatment, despite the fact that the treatment is deadly, reminiscent of an old aphorism attributed to Galen—"all who drink of this remedy are cured except in whom the disease is incurable." In the end, consistency across multiple end points is required to develop evidence of an effective treatment transcending the two analysis approaches.

Complicating the analyses of intention to treat versus per protocol are the issues of missing data. The best idea is not to have any, but often it occurs and needs to be considered in advance. For example, in the methotrexate study, a decision was made to use the multiple imputation method of intent to treat for the primary and secondary end points. Using this intention-to-treat method, while there were trends favoring methotrexate in the primary and secondary measures, none reached statistical significance. On the other hand, if the investigators had chosen either the last observation carried forward or the worst data carried forward in the intention-to-treat analyses, the study results might have yielded different conclusions (see Table 3 in Ref. 4). Using either of these alternative intent to treat analyses made most of the secondary analyses (QMG, MG-Activities of Daily Living (MG-ADL), and MG-Composite (MGC)) statistically significant in favor of the methotrexate group. This inconsistency points to key information or violations of assumptions, such as the cause of the missing data. In this example, it is likely that there was informative censoring or what is called *differential dropout*, in which the cases with the least improvement dropped out, and there were more such cases in the control arm. Thus, the a priori analysis using multiple imputations, which assumes that there is no censoring related to the outcome, may not be correct. This may not cause the study to be viewed as a positive trial, but the interpretation of the data needs to consider the totality of the evidence.

Placebo effects

The presence and magnitude of placebo effects has varied substantially across these recently published trials. Since MG trials invariably entail the addition of an experimental therapeutic to some standard of care treatment, placebo groups are rarely (if ever) entirely untreated. Conceptually, therefore, it is important to draw a distinction between therapeutic effects in the placebo group (which also receives standard of care) and placebo effects (i.e., therapeutic effects that can be attributed to administration of placebo). Here we use the term "placebo effect" to refer only to therapeutic effects that are attributable to the placebo itself. Placebo effects may have several sources: measurement error, psychological effects, or physical effects. Measurement error generates placebo effects, because we often enroll patients that exceed certain thresholds on the same scales used to measure the outcome (or scales that correlate with the outcome of interest). When individuals are measured with error, those whose errors are underestimating the true value may not enter the trial because they do not meet the eligibility hurdle, whereas those who are measured too high will be enrolled. When measured subsequently, on average the individuals who are now measured correctly on this subsequent occasion revert to their average and, thus, back toward the population average; hence the term "regression toward the mean." This phenomenon is so common in medical

research and so often confused with a therapeutic effect in single-group studies that we are going to spend a bit more time explaining this phenomenon.

Suppose we roll a die 60 times to represent observations made at screening. If we have a screening value > 3 , we say the person is eligible. As can be seen in Table 4, a fair die would give us 10 patients on average with each screening value determined by the die, and 30 of the 60 patients (those with 4, 5, or 6) would be eligible for the trial. In this scenario, baseline mean value would be $5 = (10 \times 4 + 10 \times 5 + 10 \times 6)/30$.

To define outcomes following “treatment,” let us roll the die again. Table 5 shows the expected results for the 30 eligible patients. We would expect 5 in each group by chance. The mean of the post treatment scores is $3.5 = (5 \times 1 + 5 \times 2 + 5 \times 3 + 5 \times 4 + 5 \times 5 + 5 \times 6)/30$.

The change or improvement in the “clinical score” is $5 - 3.5 = 1.5$ or a 30% reduction from baseline. This clinical improvement is due solely to random variation, and while this simple dice example does not seem biologically sound, it reflects what is happening when we select individuals with levels on a scale that we subsequently measure to quantify outcome. The biology analogy that is missing in this artificial example is the fact that we expect a correlation between the baseline and follow-up or posttreatment measurement. It can be proven that the amount of regression toward the mean is a function of the amount of regression if measures are totally independently adjusted for the correlation between the measurement (the value is 1 minus the correlation coefficient times the regression in the independent case). Stated another way, the component of the outcome measure that is not determined by biology (i.e., the correlation) will regress as if we were rolling dice in the above example.

Therefore, some improvement will occur owing to measurement error in almost any single-group study and in randomized trials with multiple groups. In randomized trials, we are protected from this nuisance because we compare the differences between groups. Both treatment groups demonstrate essentially the same amount of regression toward the mean, and thus in the difference it is subtracted out. In single-group studies, however, such changes are often taken to be treatment effects, falsely implying optimistic results when they are due to selection.

Not to be confused with a placebo effect is the presence of a therapeutic effect in the placebo group. The absence of a therapeutic effect in the placebo group in the EPITOME trial is likely attributable to the use of a lead-in design during which all patients were treated with pyridostigmine, with only those patients whose symptoms did not remit being randomized to the addition of placebo or prednisone. The absence of a true placebo effect may be attributable to the use of a compound outcome measure that combined the presence of a therapeutic effect (remission of symptoms), the absence of worsening of disease (development of generalized disease), and the absence of intolerable steroid side effects. In the MGTX trial, there were short-term improvements in both the thymectomy and non-thymectomy groups, but these were likely attributable to use of the same steroid dosing paradigm in both groups and a small regression toward the mean in both groups. If there was

a short-term placebo effect in the surgical group, it was likely masked by the short-term improvement in response to prednisone. The delayed therapeutic effect of thymectomy seems unlikely owing to a placebo effect given its durability over time as prednisone was tapered and MMS was maintained. This diminished placebo effect, therefore, is likely a function of (1) the use of a combinatorial end point and (2) the long duration of the trial (i.e., any short-term placebo effect of surgery would be lost with increasing time since thymectomy). By contrast, the placebo effect in the eculizumab trial was substantial, with 3-, 4-, and 5-point improvements in QMG score seen in 57%, 43%, and 29% of placebo-treated patients; this occurred despite a trial population consisting of treatment-refractory patients. Potential reasons for the large placebo effect in this 16-week trial include (1) patient perception of therapeutic benefit from an intravenously administered drug at a study center every 2 weeks; (2) the short duration of follow-up; (3) higher baseline QMG scores with greater day-to-day variability in degree of weakness in this treatment-resistant population; and (4) the high probability of statistical regression to the mean. Whether the placebo effect might have been reduced by use of an outcome that combined multiple independent outcomes (e.g., QMG and ADL scores) is unknown, although this conclusion is supported by preliminary data from the phase III trial of eculizumab.⁹

Conclusions

Here, we highlighted the heterogeneity of recently completed RCTs of therapeutics for patients with MG. Specifically, we highlighted the extent to which eligibility criteria effect the characteristics of the study population and how these characteristics in turn may determine the sensitivity of a trial to demonstrate a therapeutic effect. Study design, including treatment duration, strategies for steroid tapering, and selection of outcome measures, may similarly have profound effects on the ability of a trial to demonstrate a therapeutic effect. The varying magnitude of observed placebo effects is quite striking, and it is essential to understand the origins of such effects in order to make the necessary accommodations when powering future studies. Our intent has not been to criticize prior trials, but rather to reflect on their design, conduct, and outcome and to identify what lessons we can learn that might aid the design of future trials, thereby enhancing therapeutic development efforts.

References

1. Benatar M, McDermott MP, Sanders DB, et al. Efficacy of Prednisone for the Treatment of Ocular Myasthenia (EPITOME): A Randomized Controlled Trial. *Muscle & Nerve*. 2016; 53:363–369. [PubMed: 26179124]
2. Benatar M, Sanders D, Wolfe G, McDermott M, Tawil R. Design of the Efficacy of Prednisone in the Treatment of Ocular Myasthenia (EPTIOME) Trial. *Annals of the New York Academy of Science*. 2012; 1275:17–22.
3. Wolfe GI, Kaminski HJ, Aban IB, et al. Randomized Trial of Thymectomy in Myasthenia Gravis. *The New England Journal of Medicine*. 2016; 375:511–522. [PubMed: 27509100]
4. Pasnoor M, He J, Herbelin L, et al. A randomized controlled trial of methotrexate for patients with generalized myasthenia gravis. *Neurology*. 2016; 87:57–64. [PubMed: 27306628]
5. Howard JF Jr, Barohn RJ, Cutter GR, et al. A randomized, double-blind, placebo-controlled phase II study of eculizumab in patients with refractory generalized myasthenia gravis. *Muscle & Nerve*. 2013; 48:76–84. [PubMed: 23512355]

6. Aban IB, Wolfe GI, Cutter GR, et al. The MGTX experience: challenges in planning and executing an international, multicenter clinical trial. *Journal of Neuroimmunology*. 2008; 201–202:80–84.
7. Palace J, Newsom-Davis J, Lecky B. A randomized double-blind trial of prednisolone alone or with azathioprine in myasthenia gravis. Myasthenia Gravis Study Group. *Neurology*. 1998; 50:1778–1783. [PubMed: 9633727]
8. Sanders DB, Hart IK, Mantegazza R, et al. An international, phase III, randomized trial of mycophenolate mofetil in myasthenia gravis. *Neurology*. 2008; 71:400–6. [PubMed: 18434638]
9. Howard, J., O'Brien, F., Wang, JJ., Mantegazza, R. Dual Responder Analyses of Both Muscle Strength and Activities of Daily Living, Eculizumab Versus Placebo, in Refractory Generalized Myasthenia Gravis (gMG) Patients: Results from the REGAIN Study. American Academy of Neurology 2017 Annual Meeting; 2017; Boston.

Table 1

Eligibility criteria.

	EPITOME	MGTX	Methotrexate	Eculizumab
Population	Steroid-naïve OMG	GMG	Steroid-dependent GMG	Treatment-resistant GMG
Age	18	18–60 (65)	18	18–80
MGFA class	I	II, III, IV	II, III, IV	II, III, IVa
Disease duration	< 5 years	< 5 years	Any	Any
QMG criterion	1 ^a	12	None	12 and score 2 on 4 items
Prednisone	Naïve	On/off	10mg/day	Past or present
Immunosuppressive therapy	None permitted	None permitted	None with the preceding 60 days	Permitted if stable dose
Thymectomy exclusion	Any prior thymectomy	Any prior thymectomy	Within past 3 months	Within past 12 months

^aOcular QMG score.

Table 2

Study populations.

	EPITOME	MGTX	Methotrexate	Eculizumab
Average age	~ 63 years	~ 32 years	~ 67 years	~ 48 years
Disease duration	~ 6 months	~ 12 months	Not reported	~ 7 years
Baseline QMG	6 ^a	~ 12	~ 10.5	~ 18
% on prednisone	None	~ 75%	100%	~ 50%

^aOcular QMG score.

Table 3

Enrollment.

	# of sites	Enrollment duration	Target N	# screened (eligible)	N enrolled	N per site	N per month
EPTOME	6	34 months	88	145 (15)	11	1.83	0.32
MGTX	67	72 months	150	6958 (231)	126	1.88	1.75
Methotrexate	19	36 months	50	58 (50)	50	2.63	1.39
Eculizumab	16	27 months	24	30 (14)	14	0.86	0.52

Table 4

Screening for eligible patients.

Value on the roll of the die	Expected frequency	Eligible patients
1	10	Not eligible
2	10	Not eligible
3	10	Not eligible
4	10	10
5	10	10
6	10	10

Table 5

Treatment results for eligible patients.

Value on the roll of the die	Expected frequency
1	5
2	5
3	5
4	5
5	5
6	5