AMERICAN ACADEMY OF OPHTHALMOLOGY®

Editorial

Lessons from Recent Phase 3 Trial Failures: Don't Design Phase 3 Trials Based on Retrospective Subgroup Analyses from Phase 2 Trials

Philip J. Rosenfeld, MD, PhD - *Miami, Florida*
William J. Feuer, MS - *Miami, Florida*

Over the past 2 years, we've witnessed the failure of 3 phase III clinical trials designed to test new treatments for age-related macular degeneration (AMD).[1–3] Although we've seen our fair share of phase II clinical trial failures over the years, these recent phase III trial failures came as a bit of a surprise to retinal specialists who have enjoyed a remarkable string of phase III clinical trial successes over the past 12 years. Although phase III trial failures are unavoidable for a myriad of reasons in the high-risk drug approval process, we believe certain phase III clinical trial designs should be avoided to optimize the likelihood of success. Upon closer inspection of these 3 failed phase III trials, we found that they did not adhere to 2 basic rules of drug development that should provide the best chance of a phase III trial success. The best chance of success is to design a phase III clinical trial based on an unambiguously successful phase II clinical trial design. However, when the phase II trial results are equivocal or negative, there is a temptation to perform retrospective subgroup analyses of the phase II trial and then design the phase III trial based on a positive retrospective subgroup.[4] This should be avoided. As Pocock and Stone[4] write, "we find it hard to think of an example in which an apparent benefit in a subgroup in a trial with a negative outcome has led to a confirmation in a subsequent trial."

The best chance of success is to design a phase III clinical trial based on an unambiguously successful phase II clinical trial design.

In chronological order, the 3 AMD drugs that met their demise over the past 2 years were E10030 (Fovista, Ophthotech, New York, NY), lampalizumab (Genentech/Roche, South San Francisco, CA), and OHR-102 (squalamine eye drop, Ohr Pharmaceutical, New York, NY). E10030, an inhibitor of platelet-derived growth factor (PDGF), was studied in combination with inhibitors of vascular endothelial growth factor (VEGF) for the treatment of exudative AMD.[5,6] Both the preclinical and phase I studies supported the rationale for combining both anti-PDGF and anti-VEGF therapies to treat neovascularization, and the E10030 phase I results seemed to support the conclusion that classic-containing choroidal neovascularization (CNV) in AMD had the best response to combined anti-VEGF/anti-PDGF therapy. This was the basis for the phase II clinical trial design that recruited patients with eyes that had treatment-naïve classic-containing CNV.[5] At first, the phase II study results reported by Jaffe et al[6] seemed to be unambiguously positive and confirmed the previous phase I observations, but upon review of the baseline demographics in the phase II study, we found an unexplained imbalance in the baseline lesion sizes within the 3 randomized treatment groups, as shown in Table 1 of their report.[6] The investigators failed to adjust for this imbalance, and they failed to report the baseline visual acuities for these 3 groups. Based on these findings and their use of retrospective subgroup analyses to design their phase III trials as described below, we suspected that their phase III studies would fail, and they did (ClinicalTrials.gov Identifiers: NCT01944839, NCT01940900, NCT01940887).

In the phase II study of E10030, a total of 449 subjects were randomized 1:1:1 between 3 monthly treatment arms that included 0.3 mg E10030 plus ranibizumab (Lucentis, Genentech/Roche, South San Francisco, CA), 1.5 mg E10030 plus ranibizumab, and ranibizumab alone. The primary outcome was the change in visual acuity at 24 weeks. After 24 weeks, the visual acuity outcomes strongly supported a treatment benefit from the combined use of 1.5 mg E10030 and ranibizumab compared with ranibizumab alone, but these study results should have been called into question by the appearance of the significant lesion size imbalance at baseline, which became evident to all after the paper was published (Table 1).[6] Although some trialists recommend against significance tests for comparison of baseline differences between randomized groups,[7] a substantial imbalance in clinically important characteristics between groups constitutes a red flag for the interpretation of study results. We found it unusual that there was no mention of how this lesion size imbalance could have affected baseline visual acuities. Furthermore, there was no reporting of baseline visual acuities at all in this publication. Because the significance of this lesion size imbalance was not addressed, and baseline visual acuities were not reported, there was no way to assess the impact of these baseline characteristics on the differences in visual acuity outcomes between treatment groups at 24 weeks. After all, lesion size and lesion chronicity could have affected baseline visual acuity and affected the ability of these eyes to improve their visual acuity after receiving anti-VEGF therapy. Interestingly, though this imbalance was never addressed in their 2016 publication, it was reported in Ophthotech's S-1/A Securities and Exchange Commission filing in September 9, 2013.[8] On page 99 of their filing, Ophthotech stated this

https://doi.org/10.1016/j.ophtha.2018.06.002
ISSN 0161-6420/18

lesion size imbalance at baseline "had the potential to create a distortion in the analysis of mean change in the area of CNV." Although they attempted to adjust for this imbalance when analyzing the change in CNV lesion area with treatment, they did not address how this imbalance might have affected visual acuity outcomes in their published phase II results.[6]

Even if the difference in visual acuity outcomes remained significant after adjusting for this baseline lesion size imbalance, the study suffered from the fatal use of multiple retrospective subgroup analyses in their search for best responders among all the randomized patients. Although retrospective subgroup analysis is a useful tool for generating hypotheses about who might benefit most from a particular therapy, it is best to prespecify these subgroups; and even when prespecified, investigators should never rely on these retrospective outcomes as the basis for designing a subsequent phase III clinical trial.[4,7,9–13] Perhaps the best known critic of retrospective subgroup analyses is Richard Peto, a statistician at Oxford University, who has been credited with the quote, "You should always do them (retrospective subgroup analyses), but you should never believe them."[14] In short, subgroup analyses can be very helpful for confirming hypotheses and formulating new ones for testing in subsequent exploratory trials, but it is an expensive, high-risk strategy that is prone to failure when used as the basis for a phase III clinical trial.

An important flaw in the E10030 phase III clinical trial was that they used a retrospective subgroup analysis to conclude that eyes with subretinal hyper-reflective material (SHRM) had the best chance of visual acuity improvement when treated with E10030 in combination with ranibizumab. SHRM had been associated with highly exudative neovascular lesions with worse visual prognosis, and they concluded that this optical coherence tomography feature could be used as an inclusion criterion in lieu of the phase II requirement that type 2 or classic neovascularization be present.[15] As a result, they designed the phase III clinical trial to enroll patients with any exudative AMD as long as the neovascular lesion contained SHRM as defined by optical coherence tomography imaging. Thus, they changed their clinical trial design moving from the phase II trial, which enrolled classic-containing CNV, to a phase III clinical trial, which enrolled SHRM-containing CNV regardless of whether a classic component was present. Thus, they used the outcome of retrospective subgroup analyses to justify a change in their inclusion criteria when recruiting subjects into their phase III clinical trial. Even if they had performed a multivariable analysis to control for imbalances in lesion sizes and visual acuities at baseline and their phase II trial had been positive after this adjustment, the phase III clinical trial was likely to fail because they changed their enrollment criteria based on the inappropriate use of retrospective subgroup analyses as the basis for this change. They should have run a smaller phase II trial to confirm the importance of SHRM-containing CNV.

The phase III studies of lampalizumab suffered a fate similar to that of E10030.[2] Lampalizumab was studied for the treatment of geographic atrophy (GA), and its failure most likely resulted from the lack of a truly positive phase II clinical trial outcome and their reliance on a retrospective subgroup analysis when designing their phase III trial.[16] When planning a trial, designers have 3 parameters under their control to influence the proposed sample size. These are the magnitude of a clinically significant treatment effect, the statistical power to detect this effect, and the statistical significance level that is required for detection. The most desirable trial design has a reasonable effect size (not one that is unrealistically large), a high power ($\geq$80%), and a significance level or alpha error (P value) of 0.05; however, satisfying all of these conditions may result in a study that requires a large number of subjects, which is expensive. Therefore, 1 or more of these criteria may be relaxed in a phase II trial. In the lampalizumab phase II trial known as MAHALO, patients with bilateral GA secondary to AMD were randomized 1:1:1 to receive 10 mg lampalizumab every month or every other month vs. placebo. At the primary 18-month outcome, the monthly lampalizumab arm showed a 20% reduction in the mean growth of GA compared with the sham arm. The P value was 0.117, which met their prespecified significance level of $P < 0.2$. In a phase II clinical trial, investigators can prespecify the P value at whatever level they deem appropriate, which may be greater than the usual $P = 0.05$, but it means accepting a greater level of risk moving forward to a phase III trial. In MAHALO, this P value of 0.117 represented a 10% to 20% risk of an alpha error instead of the usual 5% one. Thus, this means that a phase III study, even with a sample size based on a 0.05 significance level and planned with the same design as the phase II trial, would be undertaken with an additional risk of failure. However, the investigators were lulled into complacency by the results of their genetic subgroup analysis and the scientific rationale they developed to explain the use of this subgroup. They reported carriers of the complement factor I (CFI) at-risk allele (CFI+) showed a much faster growth rate and a more dramatic treatment effect from the use of lampalizumab compared with non-carriers. However, this analysis was based on studying, at most, 31 of the 123 subjects randomized into their study. Although the lampalizumab treatment effect in the CFI+ subgroup showed a 54% reduction ($P = 0.004$) in the growth of GA compared with sham, this represented a retrospective subgroup of a prospectively randomized population, and these results should have been used as the basis for a smaller hypothesis-testing proof-of-concept study before its use in a large phase III clinical trial. Of note, the best test of whether lampalizumab reduced lesion growth more in CFI+ patients than in CFI− patients was the test of treatment-by-CFI interaction, which showed a P value of 0.014 and not the more impressive P value of 0.004 that was based on the CFI+ subgroup in isolation. Another strategy they could have used to test the validity of their genetic subgroup results would have been to study the growth rate of GA in this genetic subgroup without treatment just to see if these lesions actually grew faster. Of note, several studies have since analyzed the effect of CFI+ carrier status on the growth rate of GA, and none of these studies has confirmed the more rapid growth rate reported in the MAHALO subgroup analysis.[17–21] Not surprisingly, the CFI+ growth rate

results and the lampalizumab treatment effect results from these retrospective subgroup analyses were not reproduced in the phase III clinical trial.[2]

Based on the phase II MAHALO study with its marginal *P* value and the results of their retrospective subgroup analysis, Genentech moved forward with 2 large randomized, multicenter clinical trials investigating lampalizumab (10 mg) every month or every 6 weeks compared with sham for the treatment of GA (ClinicalTrials.gov Identifiers: NCT02247531, NCT02247479).[2] Together, the Spectri and Chroma studies enrolled more than 1800 participants in more than 275 sites in over 20 countries. The plan was to perform a subgroup analysis of the CFI+ population at 1 year and determine if the MAHALO results could be reproduced. However, at 1 year, the results could not be reproduced and trials were stopped. Although these outcomes were disappointing, they served as yet another example of how a retrospective subgroup analysis should not be used to design a phase III clinical trial, especially when it cannot be reproduced by other independent studies.

Finally, Ohr Pharmaceuticals pursued the use of topical squalamine for the treatment of exudative AMD. In their phase II clinical trial, they enrolled all types of treatment-naïve neovascular lesions, and patients were randomized to receive a ranibizumab injection at baseline followed by OHR-102 eye drops twice a day or vehicle drops twice a day (ClinicalTrials.gov Identifier: NCT01678963). All patients then received as-needed monthly injections of ranibizumab. After 36 weeks, there was a trend toward better visual acuity among the subjects receiving OHR-102, but when prespecified subgroups were analyzed, a greater visual acuity benefit was observed in eyes with classic-containing lesions and eyes with any occult CNV measuring less than 10 mm^2.[22] Based on the positive visual acuity outcomes in the subgroup with any occult CNV measuring less than 10 mm^2, a phase III trial was designed to study this retrospective subgroup (ClinicalTrials.gov Identifier: NCT02727881). Once again, a phase III clinical trial design based on a retrospective subgroup analysis failed.[3]

Two Basic Recommendations for the Design of a Phase III Clinical Trial

To minimize the risk and optimize the likelihood of success in a phase III clinical trial, we encourage all investigators to adhere to the following recommendations. One good rule to follow, which seems obvious but is frequently ignored, is the recommendation that if a phase II clinical trial is unambiguously positive, then the same core clinical trial design should be used in the subsequent phase III clinical trial. Investigators often want to improve upon the phase II primary outcome measure, and this is where they get into trouble with retrospective subgroup analyses. When trying to improve on a positive phase II study before moving into a phase III study or when trying to salvage a failed phase II study in the hope of moving directly into a phase III study, it is never a good idea to use the positive result from a retrospective subgroup analysis as the basis for a phase III clinical trial design. Although it is even possible to go directly from a phase I clinical trial to a phase III trial, it is a high-risk proposition, with failure being the likely outcome. Though we embrace subgroup analyses as a way to help design future studies and identify populations that may benefit the most from a particular therapy, the results from subgroup analyses must be validated by performing another phase II prospective, randomized clinical trial before moving into a larger phase III registration trial. Although this extra step takes more time and more money, it also avoids wasting still more money and time when the phase III study fails, and it avoids putting patients at unnecessary risk during the trial. Even though critics might argue that we are using 20/20 hindsight in offering these recommendations now that the results are known, the basic concerns and recommendations we discuss here are well known among clinical trialists across medicine.[4] After all, to the best of our knowledge, in ophthalmology there has never been a successful phase III clinical trial resulting in an approved drug that has been based on a retrospective subgroup analysis from a failed phase II clinical trial.

References

1. Ophthotech press release. Ophthotech Announces Results from Third Phase 3 Trial of Fovista® in Wet Age-Related Macular Degeneration. http://investors.ophthotech.com/news-releases/news-release-details/ophthotech-announces-results-third-phase-3-trial-fovistar-wet. Accessed May 26, 2018.
2. Holz FG, Busbee B, Chew EY, et al. Efficacy and safety of lampalizumab for geographic atrophy due to age-related macular degeneration Chroma and Spectri phase 3 randomized clinical trials. *JAMA Ophthamol*. 2018;136(6):666−677.
3. Ohr Pharmaceuticals Press release. Ohr Pharmaceutical Announces Efficacy Results from the MAKO Study in Wet-AMD. https://ir.ohrpharmaceutical.com/press-releases/detail/492/ohr-pharmaceutical-announces-efficacy-results-from-the-mako-study-in-wet-amd. Accessed May 26, 2018.
4. Pocock SJ, Stone GW. The primary outcome fails - what next? *N Engl J Med*. 2016;375(9):861−870.
5. Jaffe GJ, Eliott D, Wells JA, et al. A phase 1 study of intravitreous E10030 in combination with ranibizumab in neovascular age-related macular degeneration. *Ophthalmology*. 2016;123(1):78−85.
6. Jaffe GJ, Ciulla TA, Ciardella AP, et al. Dual antagonism of PDGF and VEGF in Neovascular age-related macular degeneration: a phase iib, multicenter, randomized controlled trial. *Ophthalmology*. 2017;124(2):224−234.
7. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet*. 2000;355(9209):1064−1069.
8. Ophthotech Corporation. Amendment no. 2 to form S-1 Registration Statement Under the Securities Act of 1933. Filed with the Securities and Exchange Commission on September 9, 2013. https://www.nasdaq.com/markets/ipos/filing.ashx?filingid=9499196. Accessed May 26, 2018.
9. Peto R. Misleading subgroup analyses in GISSI. *Am J Cardiol*. 1990;66(7):771−772.
10. Stephens R. The dangers of subgroup analysis. *Lancet Oncol*. 2001;2(1):9.
11. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in

clinical trial reporting: current practice and problems. *Stat Med*. 2002;21(19):2917−2930.
12. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet*. 2005;365(9454):176−186.
13. Peto R. Current misconception 3: that subgroup-specific trial mortality results often provide a good basis for individualising patient care. *Br J Cancer*. 2011;104(7):1057−1058.
14. Women's Health Studies Leave Questions in Place of Certainty. New York Times. https://www.nytimes.com/2006/02/19/health/womens-health-studies-leave-questions-in-place-of-certainty.html; February 19, 2006. Accessed May 26, 2018.
15. Willoughby AS, Ying GS, Toth CA, et al. Subretinal hyperreflective material in the Comparison of Age-Related Macular Degeneration Treatments Trials. *Ophthalmology*. 2015;122(9):1846−1853.e5.
16. Yaspan BL, Williams DF, Holz FG, et al. Targeting factor D of the alternative complement pathway reduces geographic atrophy progression secondary to age-related macular degeneration. *Sci Transl Med*. 2017;9(395).
17. Yehoshua Z, de Amorim Garcia Filho CA, Nunes RP, et al. Association between growth of geographic atrophy and the complement factor I locus. *Ophthalmic Surg Lasers Imaging Retina*. 2015;46(7):772−774.
18. Grassmann F, Fleckenstein M, Chew EY, et al. Clinical and genetic factors associated with progression of geographic atrophy lesions in age-related macular degeneration. *PLoS One*. 2015;10(5), e0126636.
19. Fleckenstein M, Mitchell P, Freund KB, et al. The progression of geographic atrophy secondary to age-related macular degeneration. *Ophthalmology*. 2018;125(3):369−390.
20. Rosenfeld PJ, Dugel PU, Holz FG, et al. Emixustat hydrochloride for geographic atrophy secondary to age-related macular degeneration: a randomized clinical trial. *Ophthalmology*; 2018 Apr 28. https://doi.org/10.1016/j.ophtha.2018.03.059 [Epub ahead of print].
21. Rosenfeld PJ, Berger B, Reichel E, et al. A randomized phase II study of an anti-amyloid β monoclonal antibody in geographic atrophy secondary to age-related macular degeneration. *Ophthalmology Retina*. 2018; Apr 17 [Epub ahead of print].
22. Ohr Pharmaceuticals Press release. Ohr Pharmaceutical Presents New Data From OHR-102 Phase II IMPACT Study in Wet-AMD at American Academy of Ophthalmology Annual Meeting. https://ir.ohrpharmaceutical.com/press-releases/detail/454/ohr-pharmaceutical-presents-new-data-from-ohr-102-phase-ii-impact-study-in-wet-amd-at-american-academy-of-ophthalmology-annual-meeting. Accessed May 26, 2018.

Footnotes and Financial Disclosures

Financial Disclosure(s): The authors made the following disclosures: P.J.R.: Research support − Apellis, Carl Zeiss Meditec, Genentech, Tyrogenex; Consultant − Boehringer-Ingelheim, Carl Zeiss Meditec, Chengdu Kanghong Biotech, Genentech, Healios K.K., F. Hoffmann-La Roche Ltd, Isarna Pharmaceuticals, MacRegen Inc, Ocudyne, Ocunexus Therapeutics, Tyrogenex, Unity Biotechnology; Equity interests − Apellis, Digisight, Ocudyne.

Correspondence:
Philip J. Rosenfeld, MD, PhD, Bascom Palmer Eye Institute, 900 NW 17th Street, Miami, FL 33136. E-mail: prosenfeld@med.miami.edu.