Using Delayed Start Design and Analysis to Investigate Potential Disease Modifying Effects in Alzheimer's Disease

Scott Andersen and Hong Liu-Seifert
Eli Lilly & Company, Indianapolis, IN, USA

Abstract

Alzheime's disease (AD) is the only major disease among the top 6 leading causes of death that does not have a treatment that cures, stops or even slows down the disease. Delayed start design has been proposed as an approach to demonstrate potential disease modification effect. In a delayed start design, following a standard randomized, double-blind, placebo-controlled phase, patients treated with placebo switch to the active treatment while patients treated with active treatment continue, during the delayed start period. Data from the delayed start period can be used to assess if the observed treatment effect at the end of the placebo-controlled phase, if significant, is consistent with a disease modifying effect. In particular, if the delayed start (DS) patients do not catch up with early start (ES) patients, the treatment effect may be considered consistent with a disease modifying effect. In this presentation, we will describe the statistical methodology to evaluate whether DS patients catch up with ES patients, the most important element of a delayed start design. We will also share examples of the implementations of the method in multiple real world scenarios.

Key words: Alzheimer's disease, disease modification, delayed start, early start

1. Introduction

Alzheimer's disease (AD) is the 6th leading cause of death in the US. Currently, Food and Drug Administration (FDA)-approved AD medications treat the symptoms of AD without curing, stopping or even slowing the progression of the disease. Many pharmaceutical companies (including Biogen, Eisai, Eli Lilly, Merck, and Roche) are researching potential disease modifying treatments in Phase III clinical trials. However, demonstrating whether an AD treatment is disease-modifying is not straight-forward. The scientific and regulatory communities do not have consensus on an approach to proving that a treatment modifies disease progression.

One approach is to utilize study design and associated statistical methods to show a treatment demonstrating characteristics consistent with a disease modifying effect. Delayed Start, or Randomized Start, clinical trial study designs were first proposed by Leber (Leber, 1997) as an approach to demonstrate that a treatment was disease modifying. In a Delayed Start clinical trial, all clinical trial participants receive the active treatment but are randomized to the timing of the start of active treatment: either starting early at the time of randomization or starting later. As a result, this study design has a placebo-controlled study period where clinical trial participants are randomized to active drug or placebo for sufficient duration to allow treatment separation, followed by a delayed start study period where clinical trial participants originally randomized to placebo are switched to active treatment and clinical trial participants originally randomized to active treatment remain on the active treatment. The null hypothesis in a Delayed Start trial design is that delaying the start of active treatment will have no impact on the clinical trial participants. In other words, by the end of the study both randomized groups of

clinical trial participants will have similar outcome measures. Alternatively, if a treatment were to be disease modifying, then the clinical trial participants who received the treatment later (only in the delayed-start period) would not be able to overcome this delay in treatment initiation and exhibit worse outcome measures throughout the delayed-start period.

In this paper, we will describe our approach to analyzing data from a Delayed Start clinical trial design. We will apply this approach to two sets of AD solanezumab Delayed Start studies that represent different treatment outcome scenarios. Finally, we will discuss interpretations of these delayed start trials.

2. Method

Defining Δ_1 as the treatment difference between placebo-treated patients' change from baseline and active-treated patients' change from baseline at the end of the placebo-controlled study period and Δ_2 as the treatment difference between placebo-treated patients' change from baseline and active-treated patients' change from baseline at the end of the delayed-start study period, we (Liu-Seifert etal, 2015) propose the following sequential testing of three null hypotheses for the analysis of Delayed Start clinical trial designs in order to demonstrate disease modification:

(1) H_{01}: $\Delta_1 \leq 0$ vs. H_{11}: $\Delta_1 > 0$,
(2) H_{02}: $\Delta_2 \leq 0$ vs. H_{12}: $\Delta_2 > 0$, and
(3) H_{03}: $\Delta_2 - 0.5*\Delta_1 \leq 0$ vs. H_{13}: $\Delta_2 - 0.5*\Delta_1 > 0$.

The overall logic behind the first two hypotheses is that one must show the active treatment is better than placebo in a placebo-controlled study (or study period) and at the end of the delayed start study period. The third hypothesis is needed to ensure that even if $\Delta_2 > 0$, the delayed-start patients' outcome is not coming sufficiently close to the early-start patients' outcome at the end of the delayed-start study period. It does this by testing that at least 50% of the treatment difference observed at the end of the placebo-controlled period is maintained in the delayed-start period. A symptomatic treatment would be one where H_{01} was rejected but H_{02} and H_{03} would not be rejected. Figure 1 shows this difference between a disease-modifying treatment and a symptomatic treatment.

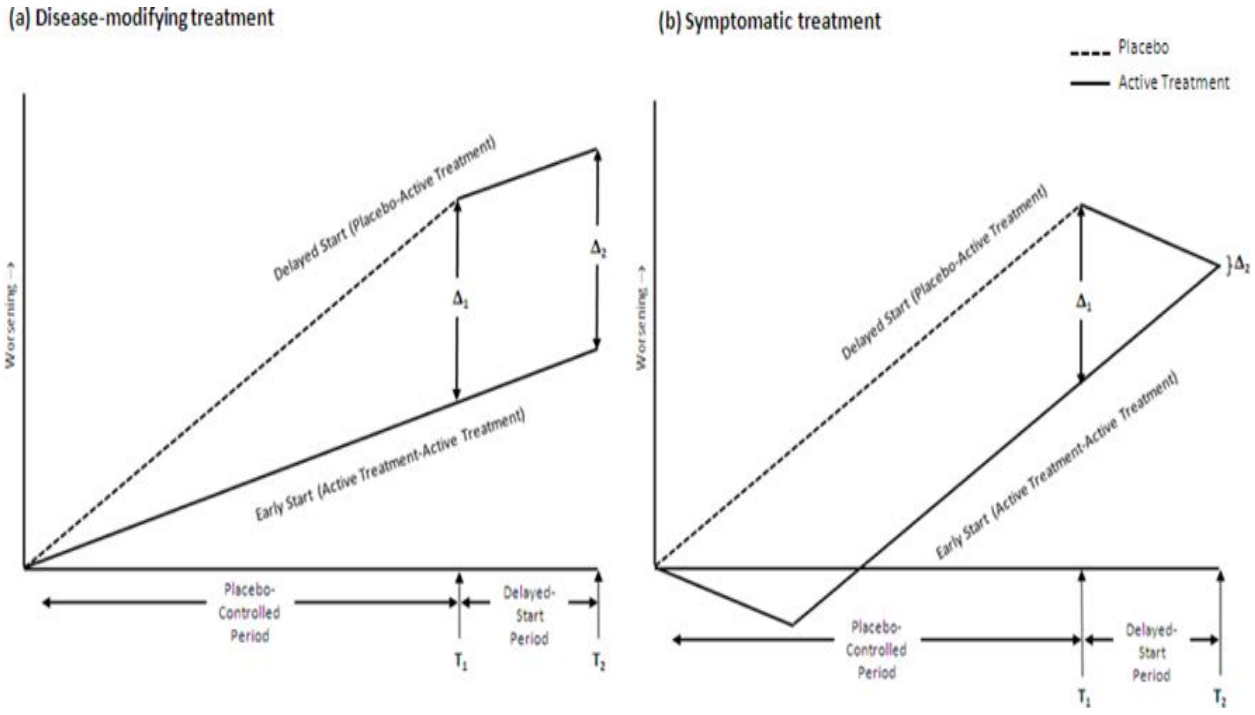

Figure 1: Disease-modifying and symptomatic treatment illustration

We propose to estimate the treatment differences from a single likelihood-based mixed effects model for repeated measures (MMRM) which includes all randomized patients and all data across the two study periods (placebo-controlled study period and delayed-start study period). It is important to use a single MMRM analysis across both study periods in order to avoid selection bias when patients discontinue early in the placebo-controlled study period because of lack of efficacy or for other reasons related to study treatment.

Several authors have proposed non-inferiority testing procedures in order to assess delayed-start data and determine if a treatment is disease modifying (Bhattaram etal 2009 and Zhang etal 2011). Our approach uses the single MMRM analysis that assesses Δ_1 and Δ_2 and tests whether a one-sided 90% confidence limit of $\Delta_2 - 0.5*\Delta_1 > 0$.

3. Application/Results

We applied the analytical methodology described above to data from patients with mild AD in Phase 3 studies (EXP, EXP2, and EXP-EXT) of solanezumab for the treatment of AD and in the most recently completed Phase 3 study of solanezumab in amyloid positive patients with mild AD (EXP3). Solanezumab is a humanized monoclonal antibody. EXP and EXP2 were 18-month, placebo-controlled studies investigating the efficacy and safety of solanezumab in patients with mild to moderate AD. EXP-EXT was an open-label extension study of up to 4 years in patients who completed EXP or EXP2. In EXP-EXT, patients who received solanezumab in EXP or EXP2 continued to receive solanezumab in EXP-EXT; patients who received placebo in EXP or EXP2 were switched to solanezumab. Patients and sites were blinded to the randomized treatment assignments throughout the entire duration of EXP, EXP2 and EXP-EXT. Thus, combining EXP, EXP2 and EXP-EXT resulted in a Delayed Start study (EXP 1/2/EXT). EXP3 was a Phase 3, double-blind study with patients randomized to solanezumab or placebo for 18 months, with an optional extension of active treatment making EXP3 as a stand-alone Delayed

Start study. The delayed-start study period included data up to 9 months following the placebo-controlled study period.

In EXP 1/2/EXT, a statistically significant difference was observed at the end of the placebo-controlled study period (Test 1) for the primary outcome, the ADAS-Cog_{14}. A statistically significant difference (Test 2) was also observed and the non-inferiority criterion was met (Test 3) at all time points, 6 months to 4 years, in the delayed-start period. Therefore, the treatment difference in the ADAS-Cog_{14} observed between the placebo and solanezumab treatment groups at the end of the placebo-controlled studies was preserved at the end of the delayed-start period within a predefined margin. However, neither Test 1, Test 2, nor Test 3 were statistically significant with regard to the primary efficacy measure, the ADAS-Cog_{14}, in EXP3.

A similar, if not as strong, finding was seen with the key secondary outcome, the ADCS-iADL, in EXP 1/2/EXT. Test 1 was statistically significant at the end of the placebo-controlled study period, and Tests 2 and 3 were statistically significant at 6 months, 1 year and 4 years of the delayed-start period. However, these significant tests were not maintained at 2 and 3 years into the delayed-start period. In EXP3, Tests 1 was statistically significant at the end of the placebo-controlled study period, and Tests 2 and 3 were both statistically significant at 6 months of the delayed-start period.

Mixed results were observed for the other two secondary endpoints in EXP 1/2/EXT. For MMSE, Test 1 was statistically significant at the end of the placebo-controlled period, but Tests 2 and 3 were not significant at 6 months of the delayed-start period and alternated between significance and insignificance throughout the rest of the delayed-start period. For CDR-SB, Test 1 was not statistically significant at the end of the placebo-controlled period, and Tests 2 and 3 alternated between significance and insignificance throughout the delayed-start period. Of the other two endpoints in EXP3, only Test 1 was statistically significant at the end of the placebo-controlled period for MMSE.

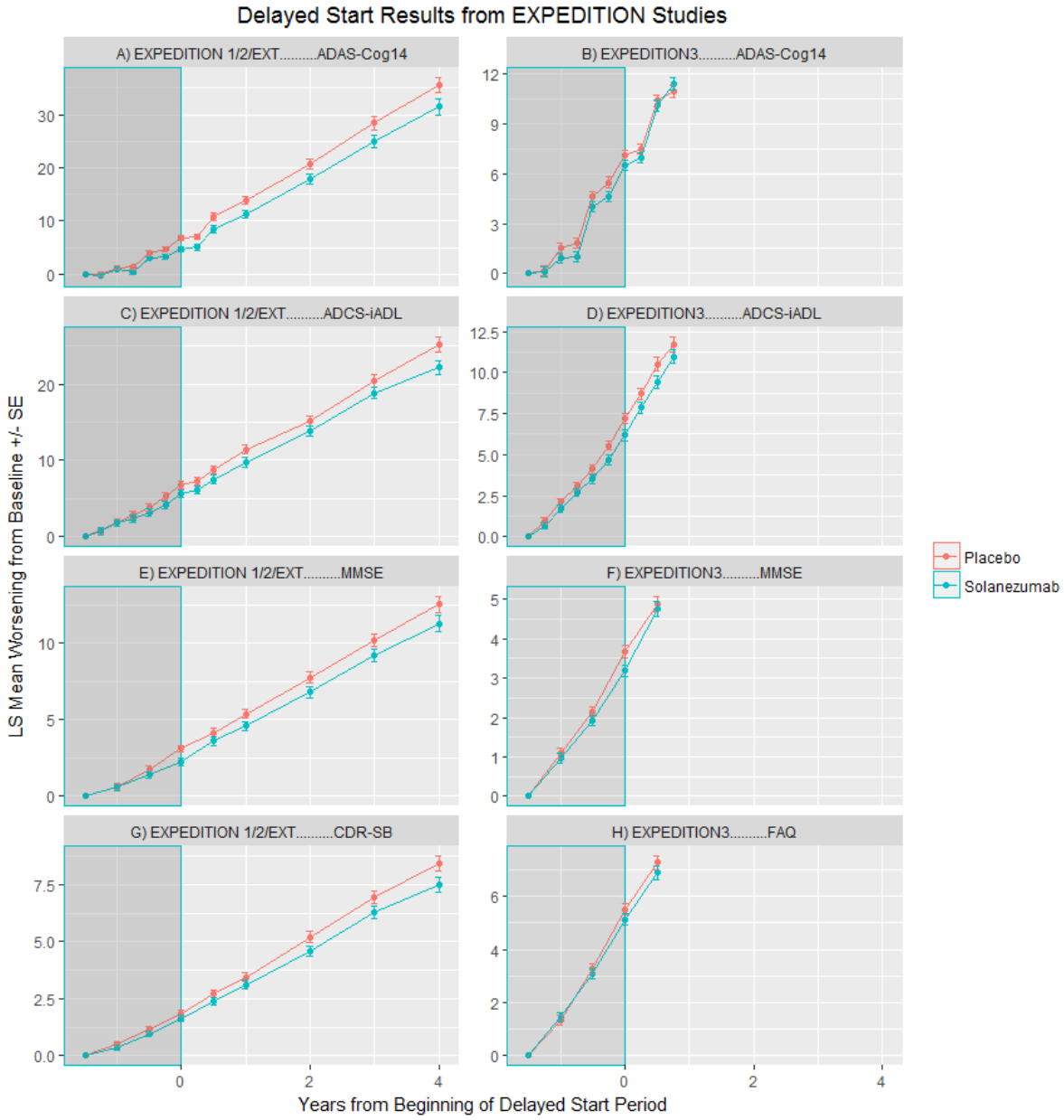

Figure 2: EXPEDITION Delayed Start Results

Table 1. Summary of EXP1, EXP2, EXP-EXT Delayed-Start Analyses

Scale Time Point	Treatment Group	Test (1) p-value	LS Mean Change from BL to Endpoint (SE)	Test (2) p-value	95% CI	Test (3) Lower Bound of CI (SE)
ADAS-Cog$_{14}$		**.002**				
6 months	DS N=441		10.76 (0.628)	**.006**	-3.90, -0.66	0.52 (0.575)
	ES N=441		8.47 (0.632)			
1 Year	DS N=389		13.92 (0.735)	**.009**	-4.53, -0.66	0.63 (0.741)
	ES N=406		11.32 (0.738)			
2 Years	DS N=303		20.77 (0.987)	**.032**	-5.57, -0.25	0.46 (1.115)
	ES N=323		17.85 (0.987)			
3 Years	DS N=228		28.41 (1.242)	**.043**	-6.87, -0.11	0.55 (1.498)
	ES N=247		24.92 (1.239)			
4 Years	DS N=185		35.50 (1.457)	**.047**	-8.04, -0.06	0.69 (1.822)
	ES N=193		31.45 (1.452)			
ADCS-iADL		**.040**				
6 months	DS N=441		-8.75 (0.552)	**.045**	0.03, 2.50	0.09 (0.465)
	ES N=443		-7.48 (0.553)			
1 Year	DS N=390		-11.43 (0.597)	**.019**	0.28, 3.05	0.35 (0.566)
	ES N=405		-9.76 (0.596)			
2 Years	DS N=308		-15.17 (0.683)	.119	-0.34, 2.96	-0.23 (0.749)
	ES N=320		-13.86 (0.678)			
3 Years	DS N=237		-20.44 (0.798)	.119	-0.41, 3.59	-0.21 (0.949)
	ES N=247		-18.85 (0.789)			
4 Years	DS N=190		-25.21 (0.936)	**.014**	0.62, 5.45	0.94 (1.179)
	ES N=195		-22.17 (0.927)			
MMSE		**.001**				
6 months	DS N=439		-4.16 (0.268)	.081	-0.07, 1.20	-0.17 (0.234)
	ES N=440		-3.59 (0.268)			
1 Year	DS N=387		-5.36 (0.296)	**.035**	0.06, 1.50	-0.02 (0.278)
	ES N=404		-4.58 (0.295)			
2 Years	DS N=303		-7.74 (0.366)	**.044**	0.03, 1.90	0.03 (0.387)
	ES N=319		-6.78 (0.363)			
3 Years	DS N=229		-10.16 (0.429)	.097	-0.17, 2.07	-0.12 (0.492)
	ES N=246		-9.21 (0.424)			
4 Years	DS N=187		-12.52 (0.507)	.071	-0.11, 2.58	0.01 (0.611)
	ES N=196		-11.28 (0.500)			
CDR-SB		.189				
6 months	DS N=440		2.71 (0.166)	.128	-0.72, 0.09	0.01 (0.149)
	ES N=436		2.40 (0.167)			
1 Year	DS N=387		3.44 (0.185)	.168	-0.79, 0.14	-0.02 (0.181)
	ES N=403		3.12 (0.185)			
2 Years	DS N=305		5.21 (0.232)	**.038**	-1.24, -1.04	0.20 (0.257)
	ES N=320		4.57 (0.231)			
3 Years	DS N=235		6.94 (0.282)	.091	-1.39, 0.10	0.10 (0.336)
	ES N=247		6.30 (0.280)			
4 Years	DS N=189		8.42 (0.330)	**.044**	-1.79, -0.02	0.27 (0.412)
	ES N=196		7.51 (0.327)			

Table 2. Summary of EXP3 Delayed-Start Analyses

Scale Time Point	Treatment Group	Test (1) p-value	LS Mean Change from BL to Endpoint (SE)	Test (2) p-value	95% CI	Test (3) Lower Bound of CI (SE)
ADAS-Cog$_{14}$		.109				
6 months	DS N=553		10.40 (0.346)	.516	-1.14, 0.57	-0.51 (0.381)
(Week 108)	ES N=553		10.11 (0.344)			
9 months	DS N=392		10.93 (0.379)	.332	-0.48, 1.42	-1.33 (0.437)
(Week 120)	ES N=403		11.40 (0.374)			
ADCS-iADL		**.019**				
6 months	DS N=551		-10.52 (0.402)	**.044**	0.03, 2.17	0.07 (0.413)
	ES N=560		-9.43 (0.400)			
9 months	DS N=395		-11.70 (0.455)	.247	-0.50, 1.94	-0.41 (0.494)
	ES N=407		-10.98 (0.451)			
FAQ		.143				
6 months	DS N=557		7.26 (0.248)	.237	-1.01, 0.25	-0.16 (0.270)
	ESN=557		6.88 (0.247)			
MMSE		**.016**				
6 months	DS N=538		-4.88 (0.189)	.607	-0.36, 0.61	-0.37 (0.204)
	ES N=548		-4.76 (0.187)			

4. Discussion

In this paper we described the characteristics of a Delayed Start clinical trial design and explained our approach to analyzing data from such a design. Our approach uses all of the data from the study in one MMRM analysis preserving the randomization integrity, avoiding the need to assume linearity, and not requiring pre-specification of a non-inferiority margin.

Studies EXP, EXP2 and EXP-EXT suggest that solanezumab may have a disease-modifying effect on the progression of AD among patients who began treatment at the mild AD stage. The data presented here include data for up to 4 years from the EXP-EXT Study for a total of a 5.5-year period, including 18 months in the placebo-controlled period and 4 years in the delayed-start period. For the ADAS-Cog_{14}, the significant treatment difference observed at the end of placebo-controlled phase was maintained at 6 months (Week 108) and subsequently at each annual assessment through Year 4, and met the non-inferiority criterion at each of those assessments through Year 4 in the delayed start period. For the ADCS-iADL, the treatment difference was maintained at 6 months, 1 year, and 4 years, meeting the non-inferiority criterion at those three time points. More variable results were observed for the Clinical Dementia Rating scale-Sum of Boxes (CDR-SB) and MMSE demonstrating evidence for significant treatment differences and meeting the non-inferiority criterion at some visits, but not in a pattern consistent with that observed for the ADAS-Cog_{14} and ADCS-iADL

Application of this delayed-start analysis to EXP3 did not yield as strong conclusions of disease modification. While the primary outcome measure was not statistically significant, the body of evidence from EXP3 indicates that the solanezumab effect was marginally significant. In light of this smaller effect size at the end of the placebo-controlled period (Test 1), the ability to detect significant differences in the delayed start period (Test 2) is reduced. Likewise, the ability to meet the non-inferiority criteria (Test 3) is also reduced.

One feature of our approach is that it does not require pre-specifying a constant/fixed non-inferiority margin. The disadvantage of pre-specifying a fixed constant non-inferiority margin was brought to light in the Attenuation of Disease Progression with Azilect Given Once-daily (ADAGIO) study. This study employed a Delayed Start design with the goal of demonstrating rasagiline modified Parkinson's disease (PD) progression. ADAGIO met all three end points in its primary objective with low dose rasagiline: (1) early start patients had statistically significantly less worsening than delayed start patients at the end of the placebo-controlled study period, (2) early start patients had statistically significantly less worsening than delayed start patients at the end of the delayed start study period, and (3) and non-inferiority relative to a pre-defined margin between the two groups with respect to the rate of change in the delayed start study period. The high dose met two of the end points but the early start patients did not show statistically significantly less worsening than the delayed start patients at the end of the delayed start study period. In fact, while non-inferiority (relative to a pre-defined margin) between the two groups with respect to rate of change was met, the delayed start treatment group numerically out-performed the early start group at the end of the delayed start study period. In other words, the delayed start patients caught up to the early start patients despite non-inferiority of rates of change being met. A US FDA advisory committee voted unanimously in 2011 against expanding the rasagiline indication for symptomatic treatment of Parkinson's disease (PD) to include wording describing disease modification based on the results of ADAGIO.

Two aspects of our approach generate much discussion: the time point to assess Δ_2 and the non-inferiority test; and the proportion of Δ_1 to preserve in the non-inferiority test. Choosing the appropriate time point to assess Δ_2 and the non-inferiority test has challenges. The time point needs to be sufficiently long to allow the delayed-start treatment group enough time to catch up to the early start treatment group if the treatment is primarily symptomatic. In contrast, if the time point is too long, the study patient attrition rate will adversely affect the power of the statistical tests. One factor that can influence choosing the

appropriate time point is the amount of time the experimental treatment remains within a patient's system. For solanezumab 6 months equated to more than 5 half-lives and was therefore considered long enough for delayed-start patients to achieve pharmacokinetic equilibrium. Additionally, the literature suggests that most symptomatic AD treatments reach peak effect in less than 6 months (Casey etal, 2010). We believe the time point of 6 months strikes the proper balance between being sufficiently long to allow the delayed-start treatment group enough time to catch up to the early start treatment group but not too long for the analysis to be adversely affected by the patient attrition rate.

The choice of non-inferiority margin can be difficult especially in field of AD which lacks specific regulatory guidance or consensus. The FDA's draft guidance on non-inferiority clinical trials cites that setting the margin needs to reflect clinical judgement with regard to how much of the control effect should be preserved by ruling out the largest clinically acceptable loss (FDA 2010 Draft Guidance). Cardiovascular outcomes studies routinely use 50% as the margin to retain 50% of the control effect in a non-inferiority study. In AD it is reasonable to believe that a treatment have both a symptomatic and disease modifying effect. Under this scenario, one would want the disease-modifying effect to outweigh the symptomatic effect. Therefore, we believe it is logical to set the margin at 50%.

The fact that our proposed method does not assume linearity of disease progression is another potential strength. Previously published approaches to analyzing data from Delayed Start clinical trials compare slopes from regression lines in the delayed start study period (Olanow etal, 2009). This assumption of linearity may be borne out in the future. However, there is no consensus in the current understanding of AD progression, and at this point linearity should not be assumed.

Another approach called the Divergence Effect Analysis (DEA) has recently been proposed (Li and Barlas, 2018). This approach is interesting as it also does not assume linearity of the disease progression and also uses one MMRM analysis across both study periods. However it does assume that the difference in progression between treatment groups over time follows a linear trend. One aspect of the DEA requires judgement and generates similar questions that our approach has generated: selecting the starting time point. The DEA starting time point corresponds to our time point to assess Δ_2. Further, the divergence approach used a constant value as the non-inferiority margin, and this constant was determined after the trial data have been observed. The importance of pre-specifying the non-inferiority margin before seeing the data and avoiding an absolute value has been discussed previously.

There are limitations to these current analyses. The EXP, EXP2 and EXP-EXT analyses are post hoc. The primary analysis of EXP was for all mild and moderate patients. The primary analysis of EXP2 did not meet its primary endpoint for ADAS-Cog14 in the mild patients. For EXP3, the delayed-start period of the study was terminated early, and this may have resulted in all the patients essentially appearing to be early "drop-outs" and consequently affecting the statistical analysis of data. A reduction in patient sample size per visit in the delayed-start period may have decreased statistical power to adequately assess the treatment difference and the non-inferiority criterion. In addition, it is important to note that the difference observed between the placebo and solanezumab treatment groups for the ADCS-iADL might be viewed as being only nominally significant since it was not corrected for multiple comparisons. In spite of these limitations, we believe our approach to analyzing AD data from a Delayed Start clinical trial design is the best method to ascertain possible disease modifying effects.

5. Conclusion

In this paper, we described our approach to analyzing data from a Delayed Start clinical trial design and applied it to two Delayed Start AD studies which have different magnitude of treatment effects and study durations. The method was useful in evaluating the putative disease modifying characteristics of a

significant treatment difference at the end of the placebo-controlled period. It was also helpful in gaining a better understanding of the overall long term outcome in the second study where a significant treatment difference was not achieved due to the small magnitude. Disease modification has proven to be an elusive goal in AD research, and we are hopeful that Delayed Start clinical trial designs coupled with our approach can be used to show an AD treatment modifies disease progression.

References

1. Bhattaram VA, Siddiqui O, Kapcala LP, Gobburu JV. Endpoints and analyses to discern disease modifying drug effects in early Parkinson's disease. AAPS J. 2009; 11: 456–464. doi: 10.1208/s12248-009-9123-2.
2. Casey DA, Antimisiaris D, O'Brien J. Drugs for Alzheimer's Disease: Are They Effective? *Pharmacy and Therapeutics*. 2010;35(4):208-211.
3. Food and Drug Administration. Center for Drug Evaluation and Research (CDER). Guidance for industry, Non-Inferiority Clinical Trials, Draft Guidance, 2010. Available: http://www.fda.gov/downloads/Drugs/Guidances/UCM202140.pdf.
4. Leber P. Slowing the progression of Alzheimer disease: methodological issues. Alzheimer Dis Assoc Disord. 1997; 11: S10–21.
5. Jianjun (David) Li & Suna Barlas (2018) Divergence Effect Analysis in Disease-Modifying Trials,Statistics in Biopharmaceutical Research, 9:4, 390-398, DOI: 10.1080/19466315.2017.1415960.
6. H. Liu-Seifert, S.W. Andersen, I. Lipkovich, K.C. Holdridge, E. Siemers. A novel approach to delayed-start analyses for demonstrating disease-modifying effects in Alzheimer's disease. PLoS ONE, 10 (2015), p. e0119632
7. Olanow CW, Rascol O, Hauser R, Feigen PD, Jankovic J, Langston W, et al. A double-blind, delayed start trial of rasagiline in Parkinson's disease. NEJM. 2009; 361: 1268–78. doi: 10.1056/NEJMoa0809335.
8. Zhang RY, Leon AC, Chuang-Stein C, Romano SJ. A new proposal for randomized start design to investigate disease-modifying therapies for Alzheimer disease. Clinical Trials. 2011; 8: 5–14. doi: 10. 1177/1740774510392255.