

**Food and Drug Administration  
Center for Drug Evaluation and Research  
Division of Neurology Products**

**Azilect (rasagiline mesylate)**

NDA 21641

Supplement 013

New Proposed Indication:

*To Slow the Clinical Progression of Parkinson's Disease*

**Peripheral and Central Nervous System Drugs  
Advisory Committee  
Background Package**

**October 17, 2011**

# **Azilect**

(rasagiline mesylate)

New Proposed Indication: To Slow the Clinical Progression of Parkinson's Disease

## **Peripheral and Central Nervous System Drugs Advisory Committee Meeting October 17, 2011 Background Package**

- I. Topics for Advisory Committee Discussion
- II. Director Memo by Russell Katz, MD  
Statistical Memo by Kun Jin, PhD: Comments on Natural History  
Staggered Start Approach
- III. Clinical Review by Leonard P. Kapcala, MD
- IV. Pharmacometrics Review by Venkatesh Atul Bhattaram, PhD
- V. Statistical Reviews by Tristan Massie, PhD and Ohidul Siddiqui PhD
- VI. Articles
  - ~Slowing the Progression of Alzheimer Disease: Methodologic Issues
  - ~Endpoints and Analyses to Discern Disease-Modifying Drug Effects in Early Parkinson's Disease
  - ~A Double-Blind, Delayed-Start Trial of Rasagiline in Parkinson's Disease

## **DISCLAIMER STATEMENT**

The attached package contains background information prepared by the Food and Drug Administration (FDA) for the panel members of the advisory committee. The FDA background package often contains assessments and/or conclusions and recommendations written by individual FDA reviewers. Such conclusions and recommendations do not necessarily represent the final position of the individual reviewers, nor do they necessarily represent the final position of the Review Division or Office. We have brought this issue to this Advisory Committee in order to gain the Committee's insights and opinions, and the background package may not include all issues relevant to the final regulatory recommendation and instead is intended to focus on issues identified by the Agency for discussion by the advisory committee. The FDA will not issue a final determination on the issues at hand until input from the advisory committee process has been considered and all reviews have been finalized. The final determination may be affected by issues not discussed at the advisory committee meeting.

### **Topics for Advisory Committee Discussion**

- 1) Does the committee believe that the randomized start design, appropriately designed and conducted, is capable of detecting a disease modifying effect for treatment of patients with Parkinson's Disease? If not, are there alternative designs that can demonstrate such an effect?
- 2) Agency reviewers have identified numerous issues related to the analyses/results of ADAGIO (Attenuation of Disease Progression with Agilect/Azilect Once Daily) and TEMPO (TVP-1012 in Early Monotherapy for Parkinson's Disease Outpatients), including:
  - a. Non-linearity of slopes, presumably related to varying early effects of treatment
  - b. Re-analyses of slopes without early data suggest parallel slopes in Phase 1 for drug and placebo
  - c. Potentially significant baseline differences in UPDRS (Unified Parkinson's Disease Rating Scale) scores between ES(early start) and DS (delayed start) patients in the Hypothesis 2 & 3 datasets, and potential biases in the analyses that compare these non-randomized groups
  - d. Differential response in men and women (primarily in ADAGIO), and baseline differences in early and delayed women starters in ADAGIO
  - e. Sponsor-conducted analyses that differed from those specified in the protocol

Please discuss the impact these issues, as well as any other issues you consider important, have on your interpretation of the studies submitted.

- 3) Does the committee find that ADAGIO provides compelling evidence that the 1 mg dose of rasagiline met the protocol specified criteria for success?
- 4) The 2 mg dose failed to show a differential effect between the early and delayed starters at the end of the study. The sponsor has offered some explanations (e.g., patients in the worst quartile of baseline UPDRS scores seemed to have a better response than other patients). Does the committee believe that the 2 mg group failed to meet the protocol specified criteria for success?
- 5) Does the committee conclude that the sponsor has provided substantial evidence of effectiveness for rasagiline as a disease-modifying treatment for patients with Parkinson's Disease?



## MEMORANDUM

DATE: September 12, 2011

FROM: Russell Katz, M.D.  
Director  
Division of Neurology Products

TO: Members and Consultants of the Peripheral and Central Nervous Systems Advisory Committee (PCNS AC)

SUBJECT: Briefing Package for the October 17, 2011 PCNS AC meeting to discuss Supplement 13 to NDA 021641, for the use of Azilect (rasagiline mesylate) as a disease modifier in patients with Parkinson's Disease

As you know, the PCNS AC will discuss Supplement 13 (S-013) to NDA 021641, for the use of Azilect (rasagiline mesylate) as a disease modifier in patients with Parkinson's Disease, at a meeting on October 17, 2011. This package contains, in addition to this overview memo, reviews of this supplement performed by several Agency reviewers, including statistical reviews by Drs. Ohidul Siddiqui, Tristan Massie, and Kun Jin, a review by Dr. Atul Bhattaram, of the Agency's Pharmacometrics group, and a clinical overview by Dr. Leonard Kapcala. In addition, we have included the following relevant articles from the medical literature:

- 1) Leber, P. Slowing the Progression of Alzheimer Disease: Methodologic Issues. *Alzheimer Disease and Associated Disorders* Vol. 11, Suppl. 5, pp.S10-S21. This article introduced the clinical trial design that the sponsor utilized to establish a disease modifying effect of rasagiline.
- 2) Bhattaram VA, Siddiqui O, Kapcala LP, Gobburu JVS. *Endpoints and Analyses to Discern Disease-Modifying Effects in Early Parkinson's Disease*. The AAPS Journal, Vol. 11, No. 3, September 2009. This article, written by FDA staff, presents further discussion about the design utilized.
- 3) Olanow CW, Rascol O, Hauser R, et al. *A Double-Blind, Delayed-Start Trial of Rasagiline in Parkinson's Disease*. *N Engl J Med* 2009;361:1268-78. This article presents the results of the primary study submitted in this supplement to support the proposed disease modifying claim.

In addition to these documents, this briefing package contains a list of "Topics for Discussion" that we would like you to address at the October meeting. You will also receive, under separate cover, a briefing package prepared by TEVA Neuroscience, Inc., the sponsor of the NDA for rasagiline.

As you know, Azilect (rasagiline mesylate) was approved in 2006 for the treatment of the signs and symptoms of idiopathic PD, both as monotherapy and as adjunctive therapy to levodopa. Rasagiline is known to act as an inhibitor of primarily monoamine oxidase inhibitor type B (MAO-B), which presumably increases extracellular dopamine in the striatum. The studies supporting approval of Azilect were designed to demonstrate an effect on the signs and symptoms of PD, but were not considered to have established that rasagiline could modify the progression of the disease process itself.

However, for various reasons, the sponsor believed that rasagiline possesses additional mechanisms that could fundamentally slow the progression of PD, not just treat the symptoms of PD. These presumed mechanisms, including neuroprotection related to rasagiline's effects on apoptotic pathways, are discussed in detail in the sponsor's briefing book. In order to establish this effect clinically, the sponsor performed a trial, ADAGIO, that purports to be adequately designed and conducted to demonstrate that rasagiline can, in fact, slow the progression of PD (in this memo, slowing the progression of PD and disease modification are used synonymously). In addition to ADAGIO, the sponsor had previously performed a trial, TEMPO, that, though primarily considered a trial that demonstrated a symptomatic effect of rasagiline and previously relied upon by the Agency to support the original approval, had some design elements that were similar to those in ADAGIO, and that could possibly be another trial that could establish a disease modifying effect of the drug. Indeed, as will be discussed later, the results of TEMPO, though not definitive, were considered to have been somewhat suggestive of a disease modifying effect. For this reason, the sponsor has submitted analyses of this trial that they consider support the data in ADAGIO.

Although, of course, the demonstration of a disease modifying effect of a drug is entirely unnecessary for approval (clearly, as rasagiline is currently approved), the identification of disease modifying effects of a drug would, obviously, be an important achievement. Unfortunately, it has not been clear what sort of data would definitively establish such an effect. Structural imaging data (e.g., MRI) has often been proposed as being a promising candidate, but it is not obvious that a change in any specific imaging modality represents a fundamental change in the underlying progression of a given disease. For example, decreased brain atrophy in a drug-treated group in a study of patients with Alzheimer's Disease compared to a control-treated group does not imply preservation of functional brain (for many reasons: what appears to be functioning brain on MRI may, in fact, represent something other than brain tissue; even preservation of anatomy does not imply that it is normally functioning anatomy, etc.). At the current time, there is no widely accepted marker of functioning brain tissue, or any other marker, in patients with PD that is accepted as being useful as a marker of drug-induced modification of the underlying disease.

However, Leber, in the article included in this package, discusses several related

clinical study designs that purport to be able to establish that a drug has an effect on the underlying progression of a disease. In one such design, patients are randomized to receive a treatment or control (usually placebo). After a period of time, at the end of which the drug has been shown to be superior to the control on a clinical outcome, patients originally on the treatment are withdrawn from treatment and treated with placebo, while patients originally treated with placebo are continued on placebo. During this second phase of the study, if patients originally on active treatment approach (and achieve) the same clinical outcomes as those patients originally assigned to (and continuing to receive) placebo, this implies that the effect of the treatment in the first phase was entirely symptomatic.

However, if patients assigned to active treatment in the first phase continue to be improved in the second phase relative to those who originally received placebo, this implies that the treatment had a component of disease modification. As he discusses, the fundamental principle underlying the interpretation of this study is that early treatment with a drug that can affect the progression of the underlying disease should result in a beneficial effect that persists after the treatment is removed, at least for some reasonable duration. This persistent effect is, in this design, represented by similar (non-inferior) slopes of the responses of the two groups in the second (withdrawal) phase. Although attractive in theory, the study is difficult to perform and interpret in practice, for several reasons, including the fact that it might need to be long (presumably, the first phase has to be long enough to allow a disease modifying effect to emerge, and the second phase has to be sufficiently long to allow any superimposed symptomatic effect to resolve off treatment); the operational definition of similar or non-inferior slopes in the second phase can be complicated (what, for example, should the non-inferiority margin be), and can result in the necessity for large sample sizes; and there may be considerable numbers of dropouts, given the required duration of the study, and, importantly, the fact that patients in the first phase who received active treatment need to be withdrawn from a presumably beneficial treatment for an extended period of time. Some of these latter considerations led to a different, but closely related, design.

In this related design, patients are randomized to treatment or placebo for an appropriate duration (just as in the randomized withdrawal design described above), but in the second phase, instead of withdrawing active treatment patients to placebo, those patients initially treated with placebo in the first phase are treated with active drug, while those initially treated with active drug continue on their treatment. Again, in the second phase, if patients switched to active drug “catch up” in the second phase, to those originally treated with, and continuing to be treated with, active drug, a symptomatic effect is concluded. If, on the other hand, patients who were originally treated with active drug continue to be improved compared to those who did not receive early treatment, a disease modifying effect is concluded. In this design (the so-called “randomized start” design), then, the underlying interpretive principle is that early treatment with a disease modifying drug should continue to provide a benefit compared to later

initiation of treatment; that is, early (longer) treatment provides a persistent benefit that cannot be achieved if treatment is delayed, presumably due to an effect on the underlying pathology that cannot be “re-captured”. This design, though perhaps less intuitively attractive than the randomized withdrawal design, has what is considered to be the distinct advantage of not having to withdraw patients from a treatment that may be beneficial, presumably with improved enrollment and retention. It is this latter design that was utilized in ADAGIO, and elements of which were included in TEMPO.

## **ADAGIO**

This was a multi-center, double-blind, study in patients with early PD that consisted of two phases: the first phase was 36 weeks long, as was the second phase. Patients were randomized into one of four groups:

- 1) rasagiline 1 mg/day in Phase 1 and Phase 2 (1 mg early start; ES)
- 2) rasagiline 2 mg/day in Phase 1 and Phase 2 (2 mg early start; ES)
- 3) placebo in Phase 1 and rasagiline 1 mg/day in Phase 2 (1 mg delayed start; DS)
- 4) placebo in Phase 1 and rasagiline 2 mg/day in Phase 2 (2 mg delayed start; DS)

The primary outcome was the Total Score of the Unified Parkinson’s Disease Rating Scale (UPDRS). The UPDRS consists of three sub-scales; they assess mental function, activities of daily living, and motor function. The scale is scored from 0-176, with higher scores indicating more severe disease. In this study, the UPDRS was assessed at weeks 12, 24, 36, 42, 48, 54, 60, 66, and 72.

The study was to be analyzed according to three hypotheses, in the following order:

- 1) **Hypothesis 1**-the contrast between the slope of drug and placebo response at Week 36 (using data from weeks 12-36; Linear Mixed Model with random intercept and slope)
- 2) **Hypothesis 2**-the contrast of scores between baseline and Week 72 (Repeated Measures)
- 3) **Hypothesis 3**-a non-inferiority analysis of the slopes of the ES and DS patients from weeks 48-72 (Linear Mixed Model with random intercept and slope)

The first hypothesis was designed to determine that a difference between treatments emerged in Phase 1, the second hypothesis was designed to determine that there was a difference between ES and DS patients at the end of the study, and the third hypothesis was to determine that an “absolute” difference between the ES and DS patients persisted during Phase 2 (that is, even though a difference between groups at the end of the study might have existed [what was

tested by Hypothesis 2], it was important to show that the two groups were not approaching each other).

In each hypothesis, each dose was compared to the relevant placebo (phase 1) or DS group (phase 2) group of the same dose. That is, for example, for Hypothesis 1, the 1 mg ES group was compared to the 1 mg placebo DS group. For Hypothesis 2, the 1 mg ES and DS groups were compared to each other, and for Hypothesis 3, the week 48-72 slopes for the 1 mg ES and DS groups were compared to each other.

However, although the primary analyses presented are as described in the paragraph above, the protocol specified analyses were to be based on all data from all groups combined. There were, however, individual significant interactions for the change from baseline at Week 72 (tested under Hypothesis 2) between site, baseline UPDRS score, and sex). For this reason, the sponsor calculated the change from baseline to Week 72 for each dose contrast using data only from that dose. It is these results that will be primarily presented here.

Although patients could enter Phase 2 before the protocol specified Week 36, for Hypotheses 2 and 3, only patients who had at least 24 weeks of treatment in Phase 1 and at least one rating in Phase 2 at Week 48 or later were included.

For the slopes analyses of Hypotheses 1 and 3, the initial values used to calculate the slopes were those that were 12 weeks into the relevant phases (Week 12 for Phase 1 and Week 48 for Phase 2). This was chosen to eliminate any effect on the analyses of obvious symptomatic effects that were considered to have occurred early; disease modifying effects were considered to have taken longer to become detectable.

For Hypothesis 3, a non-inferiority margin of slopes was chosen to be 0.15 UPDRS points/week; this margin was chosen by the sponsor, and is presumably considered to be the natural rate of decline in patients with PD. The null hypothesis for Hypothesis 3 was therefore:

$$H_0: \text{Slope}_{(ES)} - \text{Slope}_{(DS)} > 0.15$$

Given the three hierarchical hypotheses, and the fact of two dose groups, the following procedures to preserve the experiment-wise Type I error at 5% were employed (as taken from Dr. Siddiqui's description):

The Hochberg Step-Up Bonferroni method for multiple comparisons between treatment groups, in combination with the hierarchical method for the testing of the three hypotheses, were used. If Hypothesis 1 was not rejected for either one of the doses at  $\alpha=5\%$ , then the other dose was tested at  $\alpha=2.5\%$ . Each statistically significant dose (as determined by Hypothesis 1) was tested on

Hypothesis 2. Each statistically significant dose (as determined by Hypothesis 2) was then tested on Hypothesis 3.

## RESULTS

The study was performed in 14 countries. About 33% of patients were from the United States and Canada.

The following chart displays the number of patients randomized to each group:

1 mg ES	288
1 mg DS	300
2 mg ES	293
2 mg DS	295
Total	1176

The following chart displays the numbers of patients considered in the analyses for each hypothesis, with the percent of the number of patients randomized in parentheses:

	1 mg ES	1 mg DS	2 mg ES	2 mg DS
Hypothesis 1	286 (99.3)	295 (99)	290 (99)	293 (99.3)
Hypotheses 2 & 3	251 (87.2)	238 (79.9)	258 (88)	249 (84.4)

The following chart displays the percent of patients entered into various phases of the study, taken from Dr. Siddiqui's Table 1:

Phase	1 mg ES	1 mg DS	2 mg ES	2 mg DS
Entered into Phase 2				
After completing Phase 1	85%	71%	83%	73%
Early transfer to Phase 2	10%	20%	11%	20%
Discontinued in Phase 1	5%	9%	7%	7%
Entered Phase 2	100%	100%	100%	100%
Discontinued in Phase 2	13%	14%	11%	12%
Completed study	87%	86%	89%	88%

The overall mean Total UPDRS at baseline was 20.39. The mean number of days from the diagnosis of PD to study enrollment was 138 days.

The results of the various hypotheses are presented below:

### Hypothesis 1

Comparison	Slope difference	P-value
1 mg-placebo	-0.046	0.013
2 mg-placebo	-0.072	0.0001

### Hypothesis 2

Comparison	Difference at Week 72	P-value
1 mg ES-DS	-1.680	0.025
2 mg ES-DS	0.356	0.602

As pointed out by Drs. Siddiqui and Massie, if the protocol-specified analyses using all data from all dose groups was performed, the contrast for the 1 mg ES-DS analysis would yield a p-value of 0.0506, which would not be considered significant according to the protocol-specified Hochberg Bonferroni adjustment. However, as Dr. Massie notes, there are still significant interactions for the same three factors (site, baseline UPDRS, and sex) for the analyses using only the 1 mg data as the sponsor found for the combined, prospective analysis, so there are questions about the propriety of substituting this analysis for the protocol-specified analysis.

In any event, there is a clear lack of statistical significance for the 2 mg ES-DS analysis; in fact, the numerical estimate of the difference is in the “wrong” direction.

### Hypothesis 3

Although the protocol specified that the dose group(s) to be analyzed for Hypothesis 3 would be those that reached significance in Hypothesis 2 (in this case, that is, only the 1 mg ES-DS contrast), I present below the results of the analyses for both the 1 and 2 mg rasagiline groups.

Comparison	Difference in Slope deterioration (units/wk)	90% Confidence Interval
1 mg ES-DS	0.000	(-0.036, 0.036)
2 mg ES-DS	0.029	(-0.005, 0.062)

The upper limits of the confidence interval for both groups exclude the non-inferiority boundary of 0.15 units/week. This would support the conclusion that the slopes of both ES and DS groups in Phase 2 for both dose groups are non-

inferior (“equivalent”). However, as noted above, by protocol, only the 1 mg ES-DS contrast is valid.

It is of particular concern here that, although the baseline-Week 72 difference for the 2 mg ES-DS contrast is clearly non-significant, the null hypothesis for non-inferiority for the 2 mg ES-DS slopes in Phase 2 is rejected. In other words, although the Week 72 scores for the 2 mg ES and DS are not different, the analysis of the slopes of the 2 mg ES and DS in Phase 2 suggests that the slopes are “equivalent”. These findings taken together raise serious questions about the appropriateness of the non-inferiority margin chosen. That is, the non-inferiority margin permits a finding of parallel slopes in Phase 2, despite the fact that the Week 72 scores are not different (indeed, the score for the ES patients is worse than the score for the DS patients).

As Dr. Massie has noted, dropouts between Phase 1 and Phase 2 can introduce complexities into the analyses and interpretation of these trials. In particular, patients entered into, and who serve as the primary population of interest for, analyses of Phase 2, will no longer be randomized groups, making statistical comparisons treacherous.

In this regard, Dr. Massie has examined the **baseline** UPDRS scores for the intent-to-treat population (basically, the population randomized, and utilized for the analyses of the Phase 1 data) and for the primary population analyzed for Hypotheses 2 & 3.

The following chart displays these data in ADAGIO:

	Baseline UPDRS	
	ITT population	Hyp 2 & 3 population
1 mg ES	20.59	20.53
1 mg DS	20.29	19.10
P-value	0.68	0.056

Another important finding, discussed by Dr. Massie, relates to sex differences in baseline UPDRS scores in the ADAGIO and TEMPO studies.

In ADAGIO, although the Baseline-Week 72 (Hypothesis 2) analysis reveals statistical significance in the 1 mg rasagiline group, an analysis by sex reveals that the effect in this analysis arises primarily (if not entirely) from women (although the p-value for Hypothesis 2 overall for the 1 mg group is 0.025 [under an analysis that considers the data for each dose group separately], the p-value for the analysis in men is 0.98; for women, the p-value is 0.0005). The following chart displays the baseline UPDRS scores for men and women in the various treatment groups, and for the various analysis populations:

Group	ITT Baseline	Baseline for Patients In Phase 2,3 analyses
-------	--------------	--

Men

1 mg ES	20.6	20.4
1 mg DS	21.2	19.9
P-value	0.52	0.6

Women

1 mg ES	20.6	20.8
1 mg DS	18.9	17.9
P-value	0.12	0.01

Linearity

It is useful to point out that a drug that produces a disease modification effect would be expected to produce a slope that is divergent from that of a control group. This is not the same as saying that the observation of divergent slopes establishes a disease modifying effect, but it does follow that the observation of parallel, or convergent, slopes (for example, in Phase 1) speaks against a disease modifying effect.

As noted by several reviewers, an important assumption underlying the various hypotheses was that the “symptomatic” effect would, for all intents and purposes, be maximal at 12 weeks post initiation of treatment (in both phases), and that, therefore, it would be appropriate to compare the slopes of the various treatment groups using data from Week 12 on (in each period). It was presumed that any effects after that would represent the effects due to disease modification.

However, as noted by Drs. Siddiqui and Massie, in ADAGIO, the data were not linear in Phase 1 from Weeks 12-36. This can be shown to be due to considerable variability in when the “symptomatic” effect reached its maximal contribution among patients (indeed, some patients never reached a clear “break point” in Phase 1). Evaluation of Phase 1 using only Week 24-36 data are shown below, taken from Dr. Siddiqui’s Table 8:

Comparison	Change in slope	P-value
1 mg-placebo difference	0.049	0.1
2 mg-placebo difference	-0.0006	0.8

These data suggest that there are no important differences in the slopes of the placebo and rasagiline treated patients from Weeks 24-36 (of course, these slopes are linear, but they must be, given that they consist of only 2 points). This observation, though not definitive, is not consistent with a disease modification effect.

Further, as pointed out by Dr. Massie, there is also evidence that the data in Phase 3 (calculated from Week 48 on) is non-linear. According to Dr. Massie, the sponsor did not reject the assumption that the data from Phase 2 (Weeks 48-72) were linear (test for linearity, p-value=0.09). However, as noted by Dr. Massie, this was tested using the combined 1 mg and 2 mg datasets, but the separate datasets were used for testing parallelism of the separate doses in Phase 2. A test of linearity using only the 1 mg data (the 2 mg data, having failed Hypothesis 2, not being relevant for testing linearity in Phase 2) yielded a p-value of 0.04, suggesting that the data for the 1 mg group in Phase 2, from Weeks 48-72, are not linear.

#### Lack of response of 2 mg in Hypothesis 2

As a potential explanation for why the response in the 2 mg dose group differed from that in the 1 mg dose group in Hypothesis 2 in ADAGIO, the sponsor suggests that because the patients in ADAGIO had early disease, there was the possibility that a floor effect may have masked a benefit in these patients. For this reason, in their briefing book to the Committee, the sponsor presents the results of those patients with a baseline UPDRS of >25.5 and those with baseline UPDRS scores <25.5 (the top quartile of scores in this study). Those results are presented below (taken from sponsor's Tables 26 and 27, pages 81 and 82 of their briefing book):

Analyses of Patients in ADAGIO with Baseline UPDRS>25.5

Hypothesis 1

Group	Slope	Difference in slope	P-value
Placebo	0.253		
1 mg	0.144	-0.109	0.026
2 mg	0.050	-0.203	<0.0001

Hypothesis 2

Group	Change in UPDRS at Week 72	P-value
1 mg DS	5.41	
1 mg ES	2.005	0.044
2 mg DS	5.086	
2 mg ES	1.460	0.038

Hypothesis 3

Group	Slope	ES-DS
1 mg DS	0.095	
1 mg ES	0.075	-0.020
2 mg DS	0.137	
2 mg ES	0.106	-0.031

Analyses of Patients in ADAGIO with Baseline UPDRS<25.5

Hypothesis 1

Group	Slope	Difference in slope	P-value
Placebo	0.106		
1 mg	0.077	-0.028	0.13
2 mg	0.072	-0.034	0.078

Hypothesis 2

Group	Change in UPDRS at Week 72	P-value
1 mg DS	4.207	
1 mg ES	2.944	0.103
2 mg DS	2.602	
2 mg ES	3.704	0.107

Hypothesis 3

Group	Slope	ES-DS
1 mg DS	0.082	
1 mg ES	0.086	0.004
2 mg DS	0.046	
2 mg ES	0.091	0.045

However, in the sponsor’s NDA submission, they presented these data by quartile (not, as above, comparing the 4th quartile with the combined first 3 quartiles). The analyses of Hypothesis 2 for ADAGIO by quartile is presented below:

Quartile	Group	Change from Baseline UPDRS	P-value
<14	1 mg DS	4.4	0.46
	1 mg ES	3.3	
	2 mg DS	0.15	0.5
	2 mg ES	1.04	
>14<19	1 mg DS	4.45	0.22
	1 mg ES	2.6	
	2 mg DS	2.88	0.56
	2 mg ES	3.67	
>19<25.5	1 mg DS	4.12	0.99
	1 mg ES	4.10	
	2 mg DS	3.50	0.27
	2 mg ES	5.06	
>25.5	1 mg DS	4.96	0.015
	1 mg ES	0.94	
	2 mg DS	6.6	0.10
	2 mg ES	4.1	

Given that a positive change in slope indicates worsening, it is clear that there is not (a presumably expected) monotonically decreasing slope by baseline UPDRS quartile. For example, the difference in slopes between the 2 mg DS and ES groups in the third quartile moves in the “wrong” direction, favoring the DS patients.

## TEMPO

This was a multi-center, double-blind trial that, like ADAGIO, consisted of two phases: in this case, Phases 1 and 2 were each 26 weeks in duration. Patients were randomized into one of three groups:

- 1) rasagiline 1 mg/day in both Phases 1 and 2 (1 mg early start; ES)
- 2) rasagiline 2 mg/day in both Phases 1 and 2 (2 mg early start:ES)
- 3) placebo during Phase 1 and rasagiline 2 mg/day in Phase 2 (2 mg delayed start; DS)

In this trial, the primary analyses were of Phase 1 only.

Assessments in TEMPO were made at Weeks 4, 8, 14, 20, 26, 32, 42, and 52. For Hypotheses 2 & 3, patients were included who had at least one UPDRS assessment at Week 24 and at least one assessment at Week 42 or 52.

TEMPO was performed entirely in the US (88%) and Canada (12%). The following chart displays the number of patients randomized to each treatment group:

Comparison	Randomized
1 mg ES	134
2 mg ES	132
2 mg DS	138
Total	404

The results of the analyses of change from baseline in Total UPDRS score in Phase 1 are given below, taken from Dr. Bhattaram's review, Table 1:

Drug	Baseline	Change from Baseline	P-value
Placebo	24.5	3.9	
Rasagiline 1 mg	24.7	0.1	0.0001
Rasagiline 2 mg	25.9	0.7	0.0001

Although not prospectively part of the primary analysis of TEMPO, the sponsor compared the Change from Baseline at Week 52 for the 2 mg DS and the 2 mg ES patients. The p-value for this comparison was 0.024.

Given the similarity in design to ADAGIO (save for the absence of a 1 mg DS group), TEMPO was analyzed retrospectively in a manner similar to ADAGIO.

The following chart displays the numbers of patients considered in the analyses for each hypothesis, with the percent of the number of patients randomized in parentheses:

	1 mg ES	2 mg ES	2 mg DS
Hypothesis 1	125 (93.3)	123 (93.2)	135 (97.8)
Hypotheses 2 & 3	96 (71.6)	89 (67.4)	94 (68.1)

The following chart displays the percent of patients entered into various phases of the study, taken from Dr. Siddiqui's Table 1:

Phase	1 mg ES	2 mg ES	2 mg DS
Entered into Phase 2			
After completing Phase 1	83%	80%	81% (placebo)
Early transfer to Phase 2	10%	14% (placebo)	20% (placebo)
Discontinued in Phase 1	7%	6%	15% (placebo)
Entered Phase 2	100%	100%	100%
Discontinued in Phase 2	3%	5%	8%
Completed study	97%	95%	92%

The overall mean Total UPDRS at baseline was 25.03. The mean number of days from the diagnosis of PD to study enrollment was 367.75 days.

The results of the "ADAGIO" analyses for TEMPO are presented below:

### Hypothesis 1

Comparison	Slope difference	P-value
1 mg-placebo	-0.085	0.13
2 mg-placebo	-0.083	0.15

Although neither dose group should be analyzed further, given the ADAGIO protocol-specified analyses, I will present the nominal results for the subsequent analyses.

### Hypothesis 2

Comparison	Difference at Week 52	P-value
2 mg ES-DS	-1.934	0.076

### Hypothesis 3

Comparison	Difference in Slope deterioration (units/wk)	90% Confidence Interval
2 mg ES-DS	-0.100	(-2.34, 0.033)

The upper limit of the confidence interval excludes the non-inferiority boundary of 0.15 units/week. This would support the conclusion that the slopes of both ES and DS groups in Phase 2 for the 2 mg group are non-inferior (“equivalent”). However, as noted above, by protocol, this comparison is precluded by the results of testing the earlier Hypotheses.

As noted earlier, in ADAGIO, we saw that the overall significance for Hypothesis 2 (ES vs DS) in the 1 mg group seemed to be based on the results in women, and that the baseline UPDRS in women included in the analysis of Hypothesis 2 were (essentially) significantly different between the ES and DS subjects. In TEMPO, although the Baseline-Week 52 (Hypothesis 2) analysis reveals nominal statistical significance for the 2 mg rasagiline group, an analysis by sex reveals that the effect in this analysis arises primarily (if not entirely) from men (the p-value for the analysis in women is 0.7; for men, the p-value is 0.03). However, there is no statistical significance between the baseline UPDRS scores between the 2 mg ES and DS patients in men. The following chart displays the baseline UPDRS scores for men and women in the various treatment groups, and for the various analysis populations:

Group	ITT Baseline	Baseline for Patients In Phase 2,3 analyses
Men		
2 mg ES	25.1	24.3
2 mg DS	24.7	24.4
P-value	0.8	0.95
Women		
2 mg ES	26.9	26.2
1 mg DS	24.1	22.7
P-value	0.22	0.096

As discussed earlier for ADAGIO, the sponsor suggested that a floor effect in UPDRS might have been responsible for a lack of a significant difference between the 2 mg ES and DS patients in Phase 2. They contend that a showing of a greater treatment effect in patients in the 4<sup>th</sup> quartile of baseline UPDRS scores (the most impaired patients) supports this conclusion. We saw there that a comparison of the responses in all 4 quartiles suggested that the responses in the 4 quartiles did not follow a pattern that was clearly consistent with this view.

A similar phenomenon is seen in TEMPO; the relevant data are shown below (in this case, only the difference in the change in UPDRS from baseline to Week 52 between the various groups is presented):

Quartile	Group	ES-DS	P-value
<17.5	2 mg DS 1 mg ES	-3.2	0.048
	2 mg DS 2 mg ES	-3.15	0.084
>17.5<23	2 mg DS 1 mg ES	-0.39	0.85
	2 mg DS 2 mg ES	-3.7	0.06
>23<31.5	2 mg DS 1 mg ES	2.56	0.26
	2 mg DS 2 mg ES	4.48	0.04
>31.5	2 mg DS 1 mg ES	-0.256	0.92
	2 mg DS 2 mg ES	-6.05	0.028

## COMMENTS

Teva Branded Pharmaceutical Products has submitted a supplement to their approved NDA for Azilect (rasagiline mesylate) which purports to establish that Azilect has been shown, not just to be a symptomatic treatment for PD, but to modify the underlying course of PD. In support of this contention, the sponsor

has submitted the results of two controlled trials. ADAGIO, a trial specifically designed to demonstrate a disease-modifying effect, is presented as the primary source of evidence, and TEMPO, a trial not primarily designed to establish a disease-modifying effect, but which shares important design elements with ADAGIO, has been submitted as supporting evidence.

ADAGIO utilized the so-called “randomized start” design (as did, in large part, TEMPO), a design proposed by Leber in 1997 as being capable of identifying a disease-modifying, as opposed to a symptomatic, effect of a treatment. The essential principle that underlies such a study’s ability to identify a disease-modifying effect is the idea that patients in whom treatment is withheld early can never achieve the same benefit as those treated earlier. If patients in whom initiating treatment is postponed eventually “catch up” to patients in whom treatment was initiated at an earlier time, this implies that the benefit that accrued in the earlier treated patients did not fundamentally alter the course of the disease. As discussed by Leber, all Agency reviewers, the sponsor, and earlier in this memo, these ideas are operationalized in the randomized start design employed here by a requirement that the two groups (early and delayed starters) differ at the end of Phase 2, and that the slopes of the UPDRS scores in Phase 2 are “parallel” (as defined by a chosen non-inferiority margin).

The design is complicated, and presupposes that the course of the disease is fairly well understood. The course of any symptomatic effect of the drug should also be well-understood, so as to be able to adequately tease out a symptomatic effect from any disease-modifying effect. These considerations have important implications for deciding how long the various phases should be. For example, it is critical to know how long a symptomatic effect may continue to increase, so that one can tell at what point in time it is appropriate to begin to calculate the slopes of rating scores. Further, a comparison of slopes requires that the data be linear; if they are not, it may not be obvious how to analyze the data. In addition, it is critical that Phase 2 be sufficiently long in duration, so that any possible symptomatic effect can “completely” washout, and permit a fair comparison of slopes between early and delayed starters. The length of such a trial can increase the likelihood of withdrawals; because the patients in several of the critical analyses include only those who have had data in both Phases, this requires analyzing non-randomized subsets of patients, with the chance of attendant biases. And, also critically, the choice of a non-inferiority margin for judging whether or not the slopes in Phase 2 are parallel can be problematic.

Although not designed to primarily assess a disease-modifying effect of Azilect, TEMPO did include a Phase 2, and a test of the difference between 2 mg early and delayed starters suggested that those patients treated early achieved a clinical benefit superior to that achieved by those in whom treatment was initiated later. This finding suggested that rasagiline might have disease modifying effects, and ADAGIO, a study formally designed to establish such an effect (if it existed) was conducted.

In ADAGIO, as we have seen, by protocol, it appears that the 1 mg dose achieved the expected result. That is, the slope of the 1 mg group did appear to be superior to that of the placebo group in Phase 1, early starters were superior to delayed starters at the end of Week 72, and the slopes of the early and delayed starters were non-inferior. However, as noted by various Agency reviewers, it appears that the data in both Phases were **not** linear, raising questions about the appropriateness of the analyses. This non-linearity was likely due to the observation that any early (presumed symptomatic) effect did not stop increasing at Weeks 12 and 48, as was presumed at the outset that it would. Because of this, it seemed inappropriate to compare slopes including the Week 12 (in Phase 1) or the Week 48 (in Phase 2) analyses. When only data from Weeks 24 and 36 were used to calculate slopes between 1 mg and placebo, the slopes do not appear to diverge. One would expect disease modification would result in divergent slopes.

Further, as Dr. Massie has shown, the overall effect at the 1 mg dose for Hypothesis 2 seems to have been driven by the effect in women, and there is a statistically significant difference between baseline UPDRS scores between the early and late women starters. These baseline data are for those women who are included in the analysis of Hypothesis 2; this is a non-random subset of all women randomized. Although baseline differences in clinical trials can be adjusted for (because random groups are being compared), adjusting for differences in non-random subsets is problematic.

Although the results for the 1 mg dose appear to have met the protocol-specified rule for success (the above-noted issues notwithstanding), clearly this was not the case for the 2 mg dose. Although the 2 mg dose appeared to be “positive” in Phase 1, there clearly was no difference between the 2 mg early and delayed starters at the end of Phase 2 (Hypothesis 2). The sponsor suggests that this might have been related to these patients’ relatively mild symptoms, resulting in a floor effect in the UPDRS; to support this contention, they present analyses of the 4<sup>th</sup> quartile (sickest patients), which they suggest shows a larger effect in these patients. As has been noted, however, examination of the pattern of responses in all 4 quartiles suggests that the findings in the 4<sup>th</sup> quartile may not reflect a true response (putting aside the fact that these analyses were post hoc). The sponsor does acknowledge, however, that the reasons for the failure of the 2 mg dose group are not clear.

It should be noted that the protocol did not require that both doses be “positive”, or even that either **particular** dose be “positive”. That is, the protocol-specified analysis was not hierarchical with respect to dose. Either dose could have met the protocol-specified rules for success. Therefore, according to the protocol, if either the 1 mg **or** the 2 mg dose reached significance according to the 3 tested hypotheses, the study overall would have to be considered “positive”. Whether or not the 1 mg dose is considered to have achieved the protocol-specified

success criteria, given the issues previously raised, may be considered an outstanding question. But even if the 1 mg dose is considered to have met the success criteria for all 3 hypotheses, the failure of the 2 mg early starters to be distinguished from the 2 mg delayed starters at the end of the study (Hypothesis 2) raises serious questions about the interpretability of the study.

There is no obvious biological or mechanistic reason why the 2 mg dose should not be disease modifying, if the 1 mg dose is. The lack of any other explanation for why the 2 mg dose should have completely failed to meet the success criterion for Hypothesis 2 clearly raises questions about the meaning of the study, including questions about the reliability of the 1 mg findings (despite the statistical significance for the 1 mg, in this case, it is fair to ask which findings [1 or 2 mg] we should believe). The results for the 2 mg dose in ADAGIO clearly do not “replicate” the findings of TEMPO, preliminary as these latter results were. In some sense, equally troubling is the finding that, despite meeting the nominal criteria for non-inferiority of slopes in Phase 2 between the 2 mg early and delayed starters, these curves are clearly not parallel (indeed, as has been pointed out, the UPDRS score at Week 72 for the early starters is worse than that for the delayed starters). These contradictory findings (non-inferior slopes and clearly non-parallel curves) raise important questions about the appropriateness of the non-inferiority margin chosen, and about the analyses performed.

The sponsor has also submitted the results of an additional, post hoc analysis, referred to as the Natural History Staggered Start. This approach is complex and novel, and, as the sponsor notes, this approach assumes that any symptomatic effect occurs early in treatment and that this symptomatic effect persists unchanged over the course of the trial. As noted above, and by the various Agency reviewers, the assumption of an early “symptomatic” effect seems not to have been met in these trials. For this and other reasons, Agency reviewers have found the results of this analysis problematic.

In summary, Agency reviewers have identified numerous problems in the analyses of ADAGIO and TEMPO (e.g., non-linearity of slopes [presumably related to varying durations of early treatment effects among patients, undermining a basic assumption of the design and analyses], baseline differences between groups included in the analyses of Hypotheses 2 and 3, analyses presented as primary without apparent sufficient justification). However, the primary finding of concern is the fact of the clear failure of the 2 mg ES patients to have been superior to the 2 mg DS patients at Week 72 in the ADAGIO study.

Although the study can be considered “positive” entirely on the basis of positive findings for the 1 mg group (although, as discussed, the findings in this group are open to question), the failure of the 2 mg group raises serious questions about the interpretation of this study, and, therefore, about whether or not rasagiline

has been shown to have disease modifying effects. There is no obvious biological explanation for why the 2 mg dose should not be disease-modifying, if the 1 mg dose is (Dr. Kapcala, in his summary, discusses the unlikely nature of an “inverted” U-shape dose response for doses so close to each other). The sponsor’s claim that the sickest patients in the 2 mg group did, in fact, show the expected response seems less than compelling, seen in the light of the overall pattern of responses across all 4 quartiles of patients defined by baseline UPDRS scores. In addition, the nominal finding of non-inferiority of slopes in Phase 2 between the 2 mg ES and DS patients is at clear odds with the obvious lack of parallelism of these slopes, raising important questions about the sponsor’s choice of the non-inferiority margin.

The Division has never approved a treatment for disease modification in any neurodegenerative disease. This application presents the first application we have seen that presents a serious case for such a claim. The sponsor has performed one trial that incorporates the major elements of the randomized start study (and a second study incorporating numerous of these elements), a design that appears capable of permitting the identification of a disease modifying effect of a drug in the ideal case. The Agency and the company had extensive discussions prior to the conduct of the study, and came to an agreement about major design elements. We also did discuss with the sponsor the fact that, given that the results of the TEMPO study were preliminary and only suggestive, and the view that granting a disease modifying claim would have important public health consequences, a second study would need to robustly demonstrate the disease modifying effects of rasagiline as detected in a randomized start design. We continue to endorse that position. ADAGIO was designed and conducted to be that study. The results have been presented and reviewed by Agency staff, who have identified numerous problems in the analyses that raise questions about the findings in the 1 mg group. More importantly, the lack of superiority of the 2 mg ES patients compared to the 2 mg DS patients at the end of the study is troubling, and calls into question any ostensibly positive findings that might be the basis for a disease modifying claim for rasagiline. These concerns will serve as the basis for the Committee’s discussion on October 17.

One final comment.

The reviews by Agency staff included in this package all express their personal views about the data submitted. It is important to emphasize that the Division has not made a final decision on this application. Clearly, the input of the Committee will be critical to our final decision.

As always, I appreciate your efforts, both in preparing for the meeting, and, of course, at the meeting itself. I look forward to seeing you all on the 17<sup>th</sup>.



## Memorandum

**Date:** September 19, 2011

**From:** Kun Jin, PhD  
Statistical Team Leader  
Division of Biometrics 1

**To:** File, NDA 021641/S0030

**Subject:** Comments on Natural History Staggered Start Approach

In this submission, the sponsor included the results from a new analysis method that is called the natural history staggered start method (NHSS). I am not aware that this method has been published in any peer reviewed journal. This method was not discussed in the protocol that was reviewed by the agency. Drs. Massie and Siddiqui have reviewed the statistical portion of this submission independently. They discussed their concerns on the validity of the assumptions and the interpretation of the results of this model. They concluded that the results from this approach are inconclusive and the method is not appropriate for the evaluation of the disease modification claim. I generally agree with their comments and conclusions. The purpose of this memo is to provide additional arguments to support their conclusions.

The description of the proposed model is as follows (taken from the Sponsor's submission in italic font). *The model for the mean change from baseline in the clinical outcome at post-randomization time  $k$ , denoted  $t_k$  and measured in years for simplicity, can be written as a simple linear model:*

$$\mu_k = \alpha_0 + \alpha_1 x + \tau_0 y_0 + \tau_1 x y_0 + \beta_0 t_k + \beta_1 x t_k + \gamma_0 y_0 t_k + \gamma_1 x y_0 t_k,$$

*and the model for the change from baseline for an individual can be written as:*

$$\Delta y_{ik} = \alpha_0 + \alpha_1 x_i + \tau_0 y_{i0} + \tau_1 x_i y_{i0} + \beta_0 t_k + \beta_1 x_i t_k + \gamma_0 y_{i0} t_k + \gamma_1 x_i y_{i0} t_k + e_{ik}$$

*In the model above,  $\alpha_1$  and  $\tau_1$  correspond to symptomatic effects, and  $\beta_1$  and  $\gamma_1$  correspond to disease modifying effects but also include effects due to the changing magnitude of symptomatic effects over time. The parameter  $\beta_0$  is the slope of the placebo group and  $\beta_1$  is the difference in the slopes of the treatment and placebo groups for a patient of average severity at baseline.*

The graphic presentation of the model is also taken from the sponsor's NDA submission, ISE Appendix 1 page 13.



$$0 \geq \varphi \geq \beta_1. \quad (1)$$

A value of  $\varphi$  outside this range is not meaningful since it would represent the scenarios that a worsening modification effect can be estimated from symptomatic improvement scores, or an inferred disease modification effect could be more than observed total improvement.

Estimating disease modification without direct measurement is known to be a challenging problem in the field. In this model, the sponsor proposed to quantify the symptomatic component in  $\beta_1$  by

$$(\beta_0 + \varphi) \tau_1. \quad (2)$$

Then from the equation

$$\varphi = \beta_1 - (\beta_0 + \varphi) \tau_1 \quad (3)$$

$\varphi = (\beta_1 - \beta_0 \tau_1)/(1 + \tau_1)$  was derived. The NHE was derived similarly. I am not convinced why the statement for (2) is true. This is the most crucial part for the entire modeling process. To my knowledge, it is the first numerical working process to attempt to quantify exact symptomatic effect from the data containing only symptomatic improvement. From numerical point of view, I do not see why the statement for (2) is true. The sponsor has not provided a rationale for this conclusion. With this setup, small or zero value of  $\tau_1$  will tend to support claim of disease modification because the defined symptomatic component is small or does not exist. Dr. Siddiqui pointed out in his review that the estimates of  $\tau_1$  are often statistically non-significant (Table 11, Page 19).

Without further debating on the reasoning of (2), let's look at whether (3) can be used as a base equation to derive  $\varphi$ . Since (1) is a necessary condition for  $\varphi$ , the solution  $\varphi = (\beta_1 - \beta_0 \tau_1)/(1 + \tau_1)$  should at least meet the condition (1). This is not true when  $\tau_1 > 0$ .

$$\beta_0 + \varphi = \beta_0 + (\beta_1 - \beta_0 \tau_1)/(1 + \tau_1) = (\beta_1 + \beta_0)/(1 + \tau_1) < (\beta_1 + \beta_0) \quad (4)$$

leads to  $\varphi < \beta_1$  that is in contrary to (1). Here we assume  $\beta_1 + \beta_0 > 0$  as it is the case in the model. It is worth noting that in the sponsor's Table 3 in ISE Appendix 1 (Page 21 out of 68), the only claimed significant result for TEMPO 2 mg group was based on the positive estimate of  $\tau_1(0.09)$ , where NHE(-6.16) had an improvement more than observed difference  $\beta_1(-5.75)$ . Since the equation (3) fails to produce a solution to meet (1) for some  $\tau_1$ , it is less convincing that the statement for the quantity (2) is true.

Now let's look at the situation when  $\tau_1 < 0$ . With  $\tau_1$  approaching -1 from right side,  $\varphi = (\beta_1 - \beta_0 \tau_1)/(1 + \tau_1)$  goes to the positive infinity that is contrary to (1). With  $0 > \tau_1 > -1$ ,

$$\varphi \text{ in ISE Appendix 1} = (\beta_1 - \beta_0 \tau_1)/(1 + \tau_1) < 0$$

leads to the following constrained condition

$$0 > \tau_1 > \max(-1, \beta_1/\beta_0). \quad (5)$$

Here, we assume  $\beta_0 > 0$  and  $\beta_1 < 0$ , as is the case in the model. Although the solutions in both the ADAGIO and TEMPO studies meet this condition, I see no assurance that this condition would be met in other studies. The programs used to produce the estimates of the parameters are generally based on the optimization algorithm that assumes the parameters are free without constraints. Also, in both the ADAGIO and TEMPO studies, there are NHE confidence intervals that contain positive values, (Tables 2 and 3, Pages 19 and 21 in ISE Appendix 1). Without additional proof, I see no reason to rule out the possibility that the model could produce an erroneous solution with  $\varphi > 0$ .

In consideration of the above arguments, I do not support that the currently proposed NHSS model is appropriate for regulatory evaluation of the efficacy claim of the supplement.

DRAFT



# 1 Recommendations/Risk Benefit Assessment

## 1.1 Recommendation on Regulatory Action

I am deferring my final recommendation for a regulatory action until this NDA has been reviewed by the Advisory Committee planned for 10/17/11.

My review to date has identified many problems/concerns that do not support the approval of rasagiline for the claim of slowing/delaying the rate of progression of early Parkinson's Disease.

- I have concluded that the prespecified, primary analysis of the primary efficacy co-endpoints (Hypotheses # 1-3) outlined in the sponsor's SAP is not "positive" (i.e., statistically significant for testing of each sequential hypothesis) or effective for the desired claim to slow Parkinson's Disease for either rasagiline dose (1 or 2 mg daily) and have outlined my reasoning previously in detail. Whereas there is no debate that there is no evidence to suggest that the 2 mg dose is "effective" for slowing the rate of progression of Parkinson's Disease based upon results in ADAGIO, it is debatable/arguable whether the 1 mg rasagiline dose can be considered "positive"/"effective" for slowing the rate of progression of Parkinson's Disease. My negative perspective on results of the 1 mg treatment group is primarily based upon the fact that the sponsor planned to use the combined dataset in its primary analysis and later used post-hoc alternative statistical approaches to show statistically significant effects for Hypothesis # 2. The sponsor utilized post-hoc, alternative statistical approaches (which showed statistical significance) because of different patterns of change for Total UPDRS for each placebo group for 1 and 2 mg (i.e., 1 and 2 mg Delayed Start groups) in the Placebo-Controlled Phase 1 and unexpected interactions in the model and assumptions that were not met. Of significant relevance, when Dr. Massie (FDA Statistical Reviewer) applied an interaction term in the model for country (instead of study site/center because the term "country" did not suggest an interaction and "study site/center" did), the result for 1 mg for Hypothesis # 2 was not statistically significant and therefore the 1 mg dose did not statistically meet all three hypotheses/efficacy co-primary endpoints.
- The 2 mg rasagiline dose was not "effective" for slowing the progression of Parkinson's Disease. There was no suggestion that the 2 mg daily rasagiline dose slowed Parkinson's Disease progression because this dose did not statistically meet Hypothesis # 2 (treatment difference/benefit for 2 mg Early and Delayed Start groups) at the end of ADAGIO (at 72 weeks) by any primary analysis nor a host of sensitivity and supportive analyses.
- The dose response curve for ADAGIO does not make pharmacological sense because the 2 mg dose was not shown to be "effective" and arguably the 1 mg dose may have been "effective" If this effect is real, one would have to hypothesize that an inverted U-shaped

dose-response curve exists. However, this is unlikely and unrealistic for several reasons. Ordinarily, if the 1 mg dose was effective, one would expect that the 2 mg dose would be at least statistically similar (i.e., meet all three hypotheses with statistical significance as the 1 mg dose) or perhaps even numerically or statistically superior to the 1 mg dose. In other words, 2 mg might be expected to : 1) be on a similar part of the dose response curve that cannot be distinguished from the effect of 1 mg; 2) be on higher part of the dose-response curve; or 3) be on the plateau part of the dose-response curve with 1 mg. Although inverted U shaped dose-response curves can exist, they are typically demonstrated in nonclinical or in-vitro studies when one examines various doses that are very different (e.g., different by log levels) and cover a large range of dosing. Not only are inverted U shaped response not expected with merely a doubling of the dose (e.g., from 1 to 2 mg), but I am not certain if the Agency has ever observed such a drug response that it thought was real.

Negative results in ADAGIO for the 2 mg dose were also surprising because the 2 mg dose in TEMPO was the basis for the hypothesis testing of the 2 mg dose in ADAGIO. At the least, regardless of whether one considered the Early Start 2 mg dose to be statistically superior to the 2 mg Delayed Start group or not at the end of TEMPO (at 52 weeks), results for the 2 mg dose were sufficiently impressive to raise the question that 2 mg might have the potential to slow the rate of progression of Parkinson's Disease. Thus, ADAGIO did not replicate the seeming/possible benefit of early treatment with 2 mg as was suggested in TEMPO. It is not possible to seriously consider what TEMPO showed for the 1 mg dose because TEMPO's study design did not include a necessary, corresponding control group (i.e., no 1 mg Delayed Start group) for the 1 mg Early Start group. It is not feasible to draw any reasonable conclusion based upon the ad hoc comparison of the 2 mg Delayed Start group with the 1 mg Early Start group in TEMPO.

The discrepancies in the dose-response curve between ADAGIO and TEMPO are serious deficiencies that preclude granting approval of rasagiline for the desired claim of slowing the rate of Parkinson's Disease. However, even if the available data were restricted only to ADAGIO and results for TEMPO did not exist, I would still have my strong thoughts that rasagiline should not be approved at this time because of the lack of benefit of the 2 mg dose that does not make pharmacological sense.

In my opinion, I believe that the following important residual questions remain to be answered :

**Are the results of ADAGIO for the 1 mg dose a true positive or a false positive?**

**Are the results of ADAGIO for the 2 mg dose a true negative or a false negative?**

- The discrepancy in dosing results for 1 and 2 mg are not only of scientific interest but have real, important, practical implications. If rasagiline was approved for a claim for slowing the progression of early Parkinson's Disease (the only claim that can seriously be considered because an effect/benefit was not studied in advanced Parkinson's Disease), what would the

dosing section of the label recommend for the claim to obtain slowing of disease progression? If one truly believed that an inverted U-shaped dose-response curve existed and the 1 mg dose was beneficial but the 2 mg dose was not, there is a potential problem/concern for patients who take the 1 mg dose and possibly experience the higher pharmacokinetic exposure of subjects dosed with 2 mg for any known or unknown reason. In such an instance, patients could be exposed to the 1 mg dose and its safety risks and costs without the possibility of any expected benefit for slowing disease progression.

- Overall, results of TEMPO do not appear to demonstrate a clinical benefit consistent with slowing of disease progression. It is important to note that TEMPO was not prospectively planned and conducted to seek a claim for slowing disease progression. TEMPO (including the second active treatment phase) was planned primarily for assessing safety and the assessment of efficacy at the end of this phase 2 was exploratory. The sponsor has acknowledged that there was no prespecified primary efficacy endpoint for the end of the second phase of TEMPO and that the SAP was not submitted to the Agency for review and approval prior to breaking the blind to analyze results of TEMPO. Dr. Massie has also raised questions regarding assessing a benefit of the 2 mg dose group at the end of TEMPO using an LOCF approach as the sponsor has used previously.

Of significant importance, when results of TEMPO are analyzed according to the 3 sequential Hypotheses using for ADADGIO and applied to TEMPO, there is no demonstration of a benefit of 1 or 2 mg for slowing the rate of progression of Parkinson's Disease . I also recognize that TEMPO was not powered to address the three hypotheses applied for showing clinical benefit and slowing of disease progression. However, because the Agency was not given the opportunity to review the Statistical Analysis Plan (SAP) for TEMPO before the blind was broken, it is difficult to surmise what analytical approach might have been recommended by the Agency had the SAP been submitted to the Agency for review and comment before the blind was broken and before data were reviewed and analyzed by the sponsor.

- The gender effects in ADAGIO and TEMPO, which are also conflicting with each other, are not only unexpected and without explanation, but they are also extremely troubling/puzzling/disturbing with regard to the primary analysis of Hypothesis # 2 (probably the most important hypothesis for demonstrating a clinical benefit of slowing of Parkinson's Disease). The sponsor's gender subpopulation/subgroup analysis in ADAGIO showed that there was a very large treatment difference for 1 mg Early Start – 1 mg Delayed Start ( - 4.1) , which was highly statistically significant ( $p = 0.0004$ ) for females and no treatment difference (0) with a p value of approximately 1. In contrast, the gender subpopulation/subgroup analysis in TEMPO showed that there was a very large treatment difference for 1 mg Early Start – 1 mg Delayed Start ( - 2.8), which was statistically significant ( $p = 0.041$ ) for males and ONLY a very minimal/small treatment difference (- 0.227) with a p value of (0.0899).

Ordinarily gender subgroup analyses of males and females replicate the finding of efficacy demonstrated in the primary analysis of the primary efficacy endpoint. Despite the fact that the gender subgroup analyses may not be statistically significant because they are not adequately powered, they very typically at the least trend in the correct direction for supporting or replicating the primary analysis. In this application, not only are the results strikingly different with a study, but they are conflicting across studies (i.e., ADAGIO and TEMPO). It is possible that these results might merely be due to chance. However, the magnitude of the gender related benefit for the 1 mg dose in females (including substantial numbers of patients) and the very highly statistically significant p value seems so striking relative to absence of any benefit in males (including larger numbers of numbers of patients than females) and p value of about 1. These markedly opposing/conflicting results make me seriously question the overall result in all patients in ADAGIO suggesting a benefit of the 1 mg dose.

If one was to believe that these results are real, it would not be possible to ascertain which results should be believed as true! If the results of ADAGIO are believed, this drug should be approved for the desired claim only in females. If the results of TEMPO are believed, this drug should be approved for the desired claim only in males.

These discrepant gender results tend to undermine a conclusion that 1 mg rasagiline should be approved for a claim to slow Parkinson's Disease progression for men and women.

- I cannot seriously consider the long-term, open-label study results for TEMPO to be supportive that early treatment for 6 months showed a clinical benefit for Early Start vs. Delayed Start patients. The sponsor has made this suggestion. However, I dismiss these results because of several confounding factors (i.e., open-label study, lack of control, lack of randomization in patients followed, differential drop-outs rates, lack of control for other Parkinson's Disease drugs).
- One study design issue that may have confounded results observed in ADAGIO for both 1 mg treatment groups is the lack of randomization of patients at the time of entry into the active treatment Phase 2 and the imbalance of the Total UPDRS for the 1 mg dose groups that was borderline statistically significant. This is a problematic issue with this type of study and might only be addressed by randomizing patients at the end of Phase 1 and prior to entering Phase 2. This issue contributes to my view/perspective that ADAGIO results may not be reliable.

The ACTE dataset population was the key population used to test Hypotheses # 2 and # 3 in ADAGIO. More specifically, there was a noteworthy difference ( $p=0.056$ ) in the mean baseline Total UPDRS scores for the 1 mg Early Start dose group (20.53) vs. the 1 mg Delayed Start dose group (19.10) in these non-randomized ACTE datasets compared to the ITT datasets which did not suggest any noteworthy difference ( $p=0.679$ ) for the 1 mg Early Start Group (20.57) compared to the 1 mg Delayed Start group (20.25). This concerning issue is outlined in Dr. Massie's statistical review (see his review). Statistically significant

differences in other baseline characteristics were also identified in Dr. Massie's review. Considering that one desires that the populations compared for a treatment benefit are similar without noteworthy differences in baseline characteristics and that this is usually the case with randomization of treatment groups, there is a concern that these differences in the groups compared in Hypotheses # 2 and 3 may have influenced results for these hypotheses, especially Hypothesis # 2. The question is raised whether such differences in these treatment groups could have played a role in the seeming benefit of 1 mg Early Start over 1 mg Delayed Start and contributed toward a false positive Type 1 error.

It is also particularly noteworthy that the markedly positive treatment benefit of 1 mg Early Start over 1 mg Delayed Start in females in ADAGIO was also associated with a baseline difference in Total UPDRS scores of 2.9 in the ACTE datasets that were statistically significant ( $p=0.0143$ ). In contrast, Total UPDRS scores for males were similar for the 1 mg Early Start group (20.4) and 1 mg Delayed Start group (19.9) and the p-value for this difference was quite high ( $p=0.5965$ ). The question is similarly raised whether differences in these non-randomized ACTE population datasets may have contributed to the markedly positive benefit of rasagiline in females in the second, active treatment phase which was not observed in males.

- Another study design and analysis issue relates to the problem of non-linearity of slopes for the different treatment groups. Hypothesis # 1 assumed linearity of slope. However, reviews by Drs. Siddiqui and Massie (see their reviews) and analyses shown in the section for Additional Efficacy Issues/Analyses (section 6.1.10) of this review outlines concerns that slopes did not appear to be linear in the Placebo-Controlled Phase 1 despite the assumption of linearity of slopes in that phase. Consequently, slopes of different treatment groups may not have been different if one focuses on the latter part of the study and excluded early data when rasagiline would be expecting to produce its symptomatic benefit. The post-hoc analyses of Dr. Siddiqui suggest that slopes of rasagiline and placebo-treated patients may be similar/parallel and not divergent as would be expected for a drug that was slowing the progression of Parkinson's Disease. The sponsor's slope analyses also support the initial analyses conducted by Dr. Siddiqui.

The question is raised by Dr, Massie as to whether non-linearity of slopes may also be a problem in analyzing data for Hypothesis # 3 in Phase 2.

- It is also worth commenting on that the fact that the magnitude of the purported clinical benefit of an "advantage" of early treatment with 1 mg rasagiline for 9 months in ADAGIO seemed quite small because such treatment only showed a mean benefit of about 1.7 Total UPDRS units (assuming that one believes that 1 mg showed a benefit and I am not confident about this effect). It is also worth noting that it is not possible to know how long this seeming clinical advantage/benefit persists. Neither is it possible to address the issue of whether a progression slowing effect may or may not be exerted in different stages of Parkinson's Disease.

- I do not believe the sponsor's post-hoc argument that the 2 mg benefit of early treatment may have been masked because of a different symptomatic benefit in patients with more advanced disease based upon the baseline UPDRS. The various, numerous analyses outlined in the section for Additional Efficacy Issues/Analyses (section 6.1.10) clearly shows that the sponsor's hypothesis does not seem to be supported when one looks primarily at various efficacy results for TEMPO. Various analyses of TEMPO results relative to different categories of baseline Total UPDRS show the inconsistency of the sponsor's hypothesis.
- The sponsor's primary prespecified primary analysis of the co-primary efficacy endpoints used the combined dataset. This approach was based upon some assumptions. However, the sponsor did not propose testing if the assumptions were met or reasonable in its Statistical Analysis Plan (SAP) nor what alternative statistical approaches should be employed. When the sponsor broke the blind, it discovered that there appeared to be interactions with baseline Total UPDRS and study site/center. Consequently, the sponsor used an alternative statistical approach including separate datasets instead of the combined dataset.
- In considering testing of both doses (1 and 2 mg rasagiline), it is important to be mindful of the statistical approach and requirements for statistical significance. If either dose was statistically significant at  $p < 0.05$ , then the other dose would be tested at the 5 % alpha. However, if one dose was not statistically significant at  $p < 0.05$ , then the other dose would be tested at an alpha of 5 %/2 or 2.5 % and require a p-value of  $< 0.025$  for statistical significance. The order of testing of the 1 or 2 mg dose was specified.
- When the primary analysis of Hypothesis # 2 for 1 mg used the combined dataset, the p value for the difference for 1 mg Early Start – 1 mg Delayed Start at the end of the study (week 72) was -1.425 ( $p = 0.0506$ ), a not statistically significant difference because the critical alpha for statistical significant was 2.5 % ( $p = 0.025$ ). In contrast, when the primary analysis of Hypothesis # 2 for 1 mg used the separate datasets, the p value for the difference for 1 mg Early Start – 1 mg Delayed Start at the end of the study (week 72) was -1.680 ( $p = 0.025$ ), a p value that just met the minimal criterion for statistical significant difference because the critical alpha for statistical significant was 2.5 % ( $p = 0.025$ ).
- The sponsor had noted post-hoc (after breaking the blind and conducting various analyses) that alternative statistical approaches for the primary analysis of the primary efficacy endpoints could employed by using separate datasets instead of the combined dataset or by adding interaction terms to the model. The sponsor chose the alternative statistical approaches of analyzing the separate datasets and adding the interaction terms of baseline UPDRS and center\*dose to the model and found that both approaches showed statistically significant effects for the 1 mg dose.

One Statistical Reviewer, Dr. Siddiqui (see his Review), thought that the sponsor's post-hoc approach was reasonable given the interactions noted. Another Statistical Reviewer, Dr. Tristan Massie (see his Review), thought that the prespecified primary analysis called for analyzing with the combined dataset. In addition, Dr. Massie noted that an interaction still

persisted if one used the separate datasets and that one way of eliminating the significant study site/center interaction was by pooling sites by country. When Dr. Massie applied a model using country instead of study cite/center, the difference at the end of the study (week 72) for 1 mg Early Start – 1 mg Delayed Start was – 1.36 and the p value was not statistically significant ( $p = 0.0873$  for the separate dataset;  $p=0.1178$  for the combined dataset) regardless of whether the separate or combined dataset was used.

Because the sponsor did not propose and prespecify in its SAP an alternative statistical approach/plan for the primary analysis if assumptions were not met, it is not possible to know whether the alternative statistical approach/plan proposed by the sponsor would necessarily have been the same alternative statistical approach had an alternative analysis been prespecified in the SAP if assumptions were not met. significance for results. It is possible that the sponsor evaluated various statistical approaches post-hoc after breaking of the blind and then proposed alternative statistical approaches/plans that showed the results to be “positive” (i.e., statistically significant).

It is unfortunate that the sponsor did not prospectively propose to test assumptions and correspondingly propose alternative statistical approaches/plans if the assumptions in its SAP were not met.

- Based upon events that I have described above here, I conclude that the prespecified primary statistical analysis of the primary efficacy endpoints was not positive for showing results suggesting that rasagiline treatment with either 1 or 2 mg daily rasagiline slowed/delayed the rate of progression of Parkinson's Disease.
- There is no replication of the effect/clinical benefit of slowing the rate of progression of Parkinson's Disease in two studies, an AGENCY standard frequently applied for approving a drug for a new claim. The Agency has previously gone on record as having told the sponsor two positive trials would be needed for the claim to slow/delay Parkinson's Disease progression. The following is a quoted excerpt from an Agency telecon meeting (12/22/04) with the sponsor. The question had been raised about whether one or two studies would be needed for the desired claim.

*“The Number of Required Studies*

- *Ordinarily 2 trials are required to support efficacy.*
- *The TEMPO study post hoc analysis may not be sufficient for review because it is not the primary analysis. If the next study is robustly positive, then the TEMPO study may provide supporting evidence.”*

It seems clearly the case that results from ADAGIO or TEMPO cannot be considered robust. Neither can I affirm the view that results from TEMPO replicate results from ADAGIO because of the conflicting results with regard to dose in each study and the fact that TEMPO cannot be considered as having demonstrated slowing of disease progression.

Despite the fact that many sponsors are interested in such a drug claim to slow the rate of progression of Parkinson's Disease and the published literature supports the view that a drug that slows/delays the rate of progression of Parkinson's Disease is a goal of utmost importance, the Agency and DNP has never reviewed an NDA application seeking such a claim. Neither are there any drugs that have been approved for slowing any neurodegenerative process. I suggest that the treatment claim desired by the sponsor for slowing the rate of progression of Parkinson's Disease could be considered as the "Holy Grail" for treatment claims in Parkinson's Disease, It is also worth noting that the sponsor seeks a broad claim ("AZILECT is indicated for the treatment of patients with idiopathic Parkinson's disease to slow clinical progression and treat the signs and symptoms of Parkinson's disease as initial monotherapy and as adjunct therapy to levodopa.") for treating all Parkinson's Disease patients (i.e., "early" and "advanced" Parkinson's Disease despite the fact that the sponsor only studied patients with "early" Parkinson's Disease with monotherapy. It is not appropriate to draw any conclusions about slowing the rate of progression of patients with advanced Parkinson's Disease when the sponsor only studied patients with early Parkinson's Disease for an effect on progression. Consequently, granting an approval by the Agency for a claim that a certain drug slows Parkinson's Disease progression would be precedent setting and should clearly be based upon robust results. Considering all the data in reviewed in this NDA to date, it seems that results are not robust and that it is difficult to point to adequate data/evidence supporting the approval of this drug, rasagiline/Azilect for the desired claim of slowing clinical progression of Parkinson's Disease.

## **1.2 Risk Benefit Assessment**

There is no demonstrated benefit of rasagiline for slowing the rate of progression of Parkinson's Disease.

## **1.3 Recommendations for Postmarket Risk Evaluation and Mitigation Strategies**

None

## **1.4 Recommendations for Postmarket Requirements and Commitments**

None



## PHARMACOMETRICS REVIEW

NDA	21-641
Submission Date	12/23/2010
Generic Name	Rasagiline
Brand Name	AZILECT®
Dose & Dosage Form	Once daily, 1 mg Tablets
Indication	For the treatment of patients with idiopathic Parkinson's disease to slow clinical progression and treat the signs and symptoms of Parkinson's disease as initial monotherapy and as adjunct therapy to levodopa.
Sponsor	TEVA Neuroscience
Submission Type	Standard
OCP Division	Division of Pharmacometrics
OCP Reviewers	Venakatesh Atul Bhattaram, PhD
Pharmacometrics Team Leader	Yaning Wang, PhD

## Table of Contents

1. Executive Summary .....	2
2. Recommendations .....	2
3. Background .....	2
4. Key Questions .....	6
4.1 Analysis of change from baseline in total UPDRS score at 72 weeks in ADAGIO (Delayed Start Design) showed that 1 mg has potential disease modifying effects while 2 mg does not. Are there any issues related to delayed start design for this finding? .....	6
4.2 Based on the observation that the assumptions in delayed start design are not the reasons for lack of disease modifying effects of 2 mg, are there any other reasons for these findings? .....	10

## 1. Executive Summary

TEVA Neuroscience evaluated rasagiline for its ability to modify the course of Parkinson's disease in two delayed-start design clinical trials (ADAGIO, TEMPO). The trials, combined, enrolled about 1500 patients who were diagnosed early in their onset of disease. The duration of TEMPO and ADAGIO clinical trials were 52 and 72 weeks respectively. Patients were randomized to four groups (1 mg delayed start group, 1 mg early start group, 2 mg delayed start group, 2 mg early start group) and three groups (2 mg delayed start group, 1 mg delayed start group, 1 mg early start group) in ADAGIO and TEMPO trial respectively.

Patients randomized to delayed start group received placebo for 0-36 weeks followed by rasagiline during 36-72 weeks in the ADAGIO trial. Patients randomized to early start group received rasagiline during 0-72 weeks in ADAGIO trial. Patients randomized to delayed start group received placebo for 0-26 weeks followed by rasagiline during 26-52 weeks in the TEMPO trial. Patients randomized to early start group received rasagiline for 0-52 weeks. Assessment of disease modification effect of rasagiline in ADAGIO and TEMPO was based on differences in slopes of disease progression in placebo vs treatment groups (0-36 weeks in ADAGIO; 0-26 weeks in TEMPO) followed by differences in early and delayed start groups at the end of the study (72 weeks; ADAGIO, 52 weeks; TEMPO). The maintenance of disease modifying effect was also analyzed by testing for parallelism of data between 48-72 weeks in ADAGIO and 32-52 weeks in TEMPO trial. The data from ADAGIO suggest that 1 mg has disease modifying effects while 2 mg does not have disease modifying effects. The reasons for these findings are not clear. There are differences in changes in total UPDRS scores between 1 and 2 mg delayed start groups during 0-36 weeks for which reasons are not clear.

Overall, the data from ADAGIO and TEMPO clinical trials suggested that the effects of rasagiline on slowing the progression of Parkinson's disease are not conclusive.

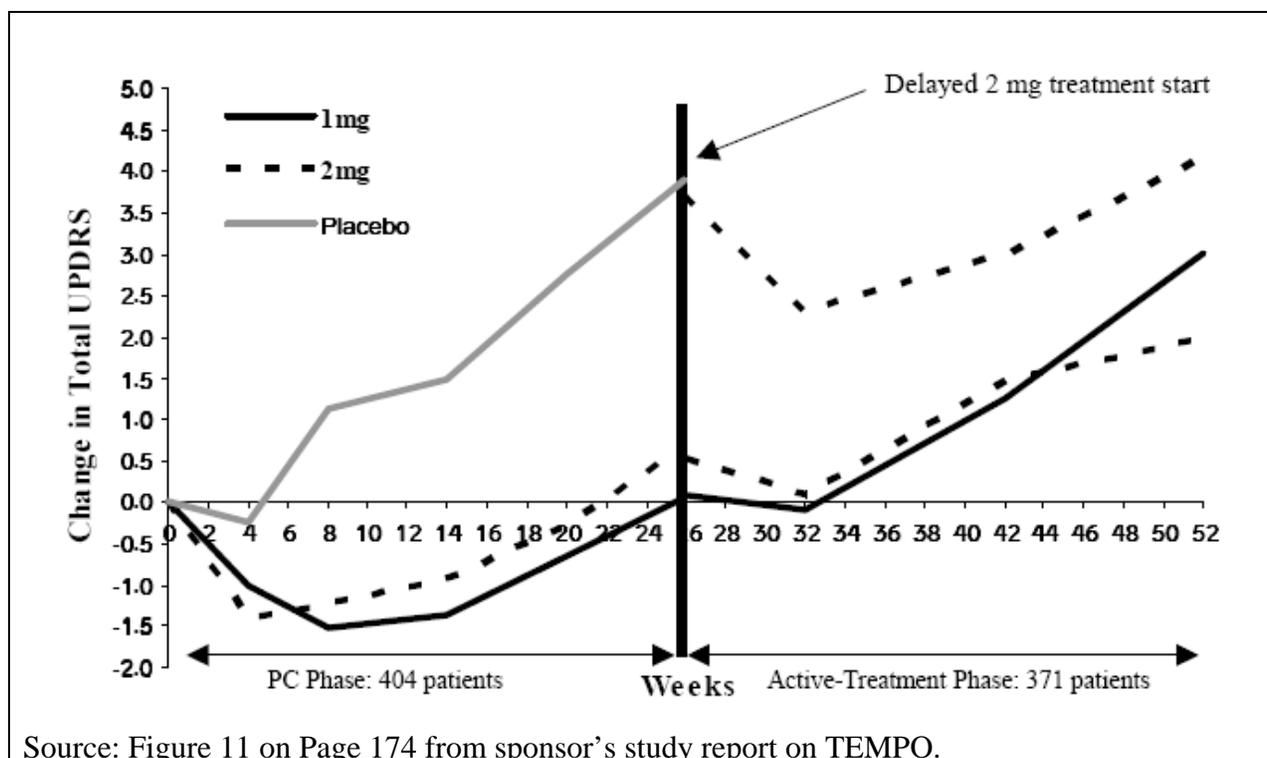
## 2. Recommendations

The pharmacometrics division has reviewed the data submitted by sponsor. Due to unexplained reasons for lack of disease modifying effect of 2 mg dose group, we recommend that the sponsor conduct a new clinical trial.

## 3. Background

AZILECT® (Rasagiline) is currently indicated for the treatment of the signs and symptoms of idiopathic Parkinson's disease (PD) as initial monotherapy and as adjunct therapy to levodopa. For the approval of rasagiline (in 2006) as initial monotherapy in PD, sponsor conducted a clinical trial (TEMPO, Multicenter, randomized, double-blind, parallel group) to assess the efficacy, tolerability and safety of two doses (1 mg, 2 mg) of rasagiline in early PD patients (N=404) not treated with levodopa. The double-blind design was maintained during the entire study; a 26-week placebo-controlled treatment was followed by a 26-week active treatment. Figure 1 shows the mean change from baseline in total UPDRS (LOCF) in TEMPO.

Figure 1. Mean Change from Baseline in Total UPDRS (LOCF) in TEMPO



For approval of rasagiline as monotherapy in PD, the data collected at 26 weeks in TEMPO trial was included in the primary statistical analysis. **Table 1** displays the results of TEMPO trial.

Table 1. Change in Total UPDRS Score at 26 weeks in Parkinson’s Disease Patients not on Dopaminergic Therapy (Monotherapy) in TEMPO trial

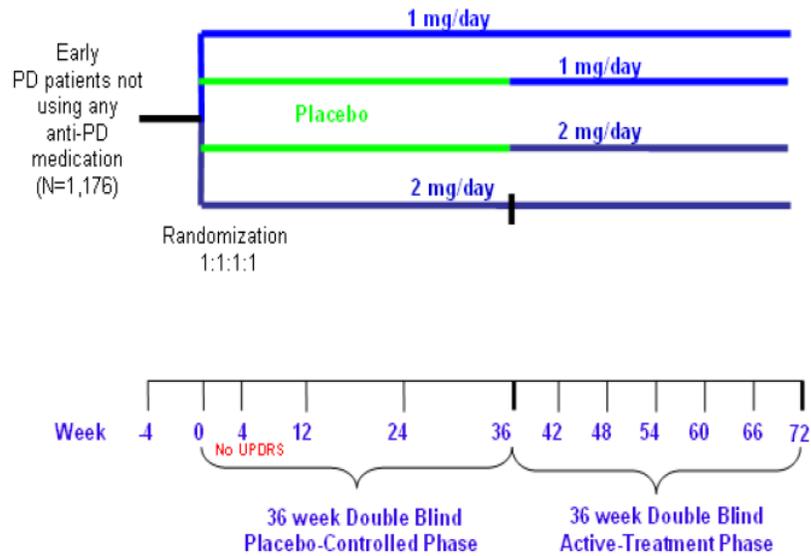
Primary Measure of Effectiveness: Change in total UPDRS score			
	Baseline score	Change from baseline to termination score	p-value vs. placebo
Placebo	24.5	3.9	---
1.0 mg/day	24.7	0.1	0.0001
2.0 mg/day	25.9	0.7	0.0001

Source: Table 3 in approved AZILECT® label

The sponsor also analyzed the differences (in change in total UPDRS score) between patients treated with 2 mg for 52 weeks (Early-Start) and patients treated with placebo for 26 weeks followed by 2 mg for additional 26 weeks (Delayed-Start). The difference gained between the placebo and 2 mg group at the end of the placebo-controlled phase (Week 26) was sustained for additional 26 weeks although both groups were treated with 2 mg rasagiline during that period. This difference was statistically significant (p=0.024). The findings from TEMPO study indicated that rasagiline might have disease-modifying effects. In this current application, sponsor submitted data from ADAGIO, a clinical study conducted using delayed-start design

principles, that evaluated the disease modifying effects of rasagiline. Figure 2 shows the study design of ADAGIO conducted in 1100 patients across 14 countries.

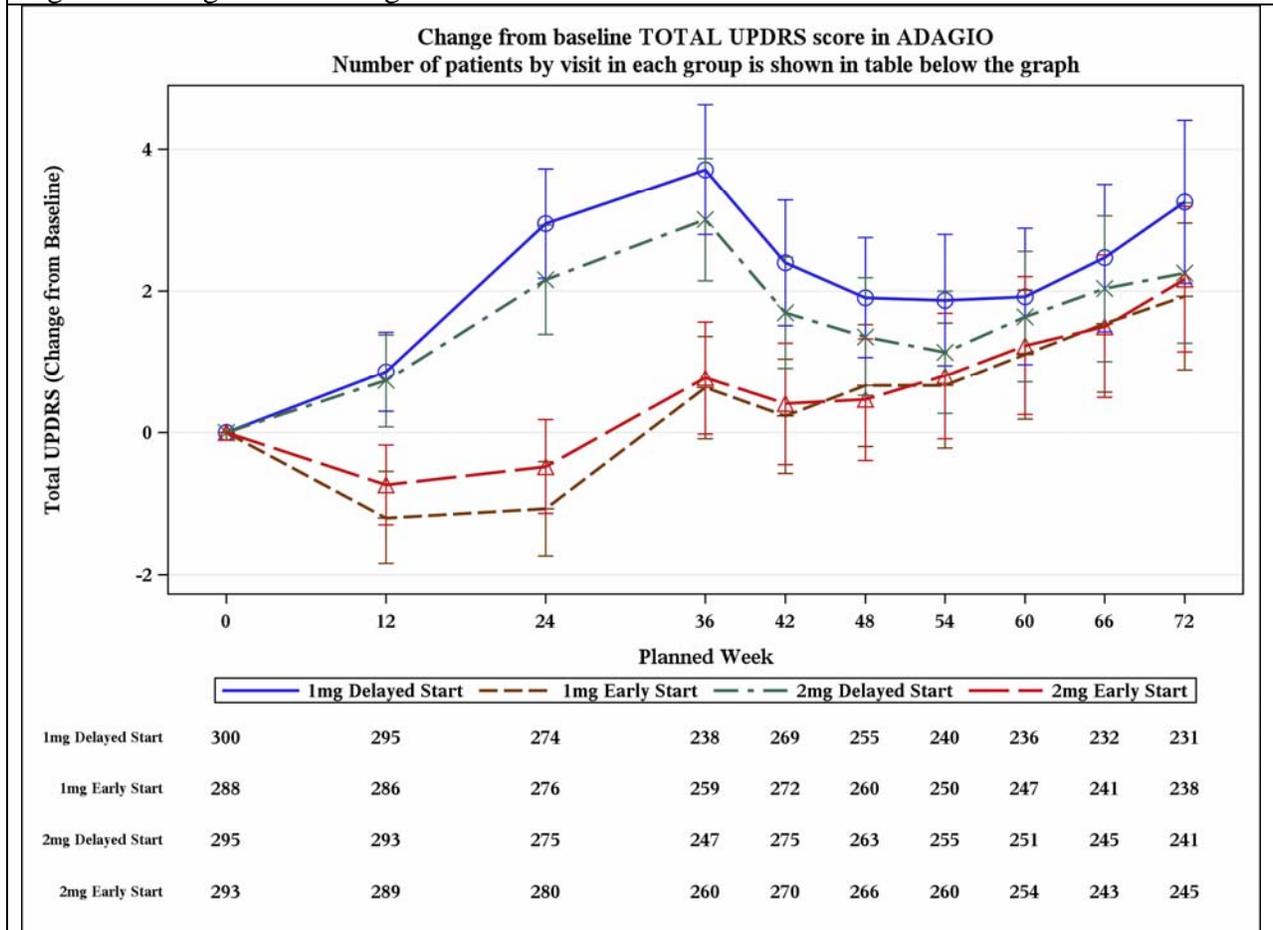
Figure 2. ADAGIO study design



Source: Figure on page 58 from sponsor's study report on ADAGIO

The results of ADAGIO trial are shown in Figure 3.

Figure 3. Longitudinal change from baseline in Total UPDRS score in ADAGIO.



For approval of rasagiline as a disease modifying drug, the sponsor was recommended to analyze the data from ADAGIO using three endpoints (as shown in Table 2) in sequence.

Table 2. Statistical analysis methodology for ADAGIO study

<b>Placebo Control Phase (0-36 weeks)</b>
Is the rate of change (slope) in total UPDRS score different between placebo (delayed start) and treatment (early start) groups between 12-36 weeks?
<b>Active Control Phase (36-72 weeks)</b>
<i>If the slope in treatment group is shallower than placebo group then:</i>
Is the mean change in total UPDRS score at 72 weeks lower in patients who started early on treatment when compared to those taking placebo for 36 weeks followed by treatment for 36-72 weeks?
If there a consistent difference in mean change in total UPDRS score between early and delay start groups from 48 weeks to 72 weeks? (Test for parallelism)

Overall, the results based on assumptions of linearity and similar symptomatic effects in patients who start treatment at 0 vs 36 weeks, the data suggest that 1 mg rasagiline has disease modifying effects. Due to lack of statistical difference between early and delayed start groups at 72 weeks

for 2 mg rasagiline, it is concluded that 2 mg does not have disease modifying effects (for further details refer to statistics review by Dr Ohid Siddiqui and Dr Massie Tristan, Office of Biostatistics).

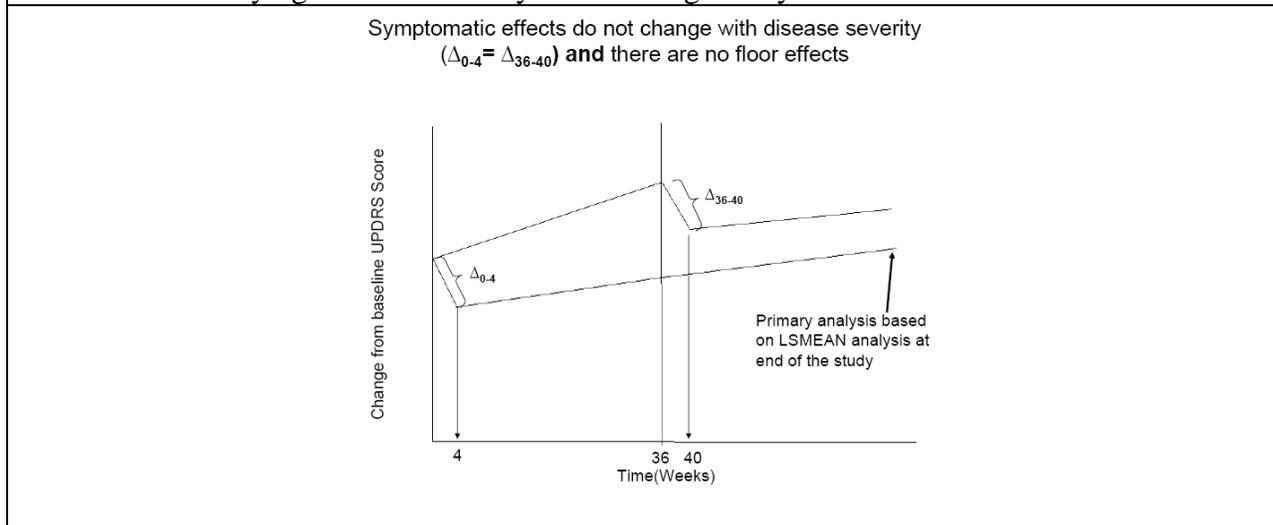
The aim of this review is to highlight the assumptions of delayed start design and understand the reasons for the observed differences in treatment effects between 1 and 2 mg dose groups in ADAGIO trial.

## 4. Key Questions

### 4.1 Analysis of change from baseline in total UPDRS score at 72 weeks in ADAGIO (Delayed Start Design) showed that 1 mg has potential disease modifying effects while 2 mg does not. Are there any issues related to delayed start design for this finding?

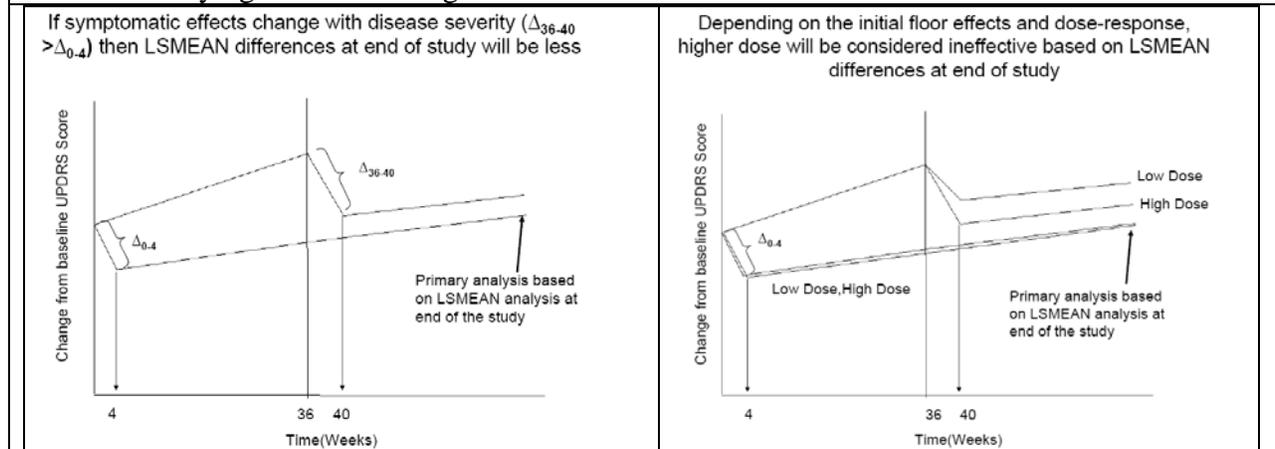
Figure 4 shows the time course of change from baseline in total UPDRS score for a hypothetical drug that has a mixture of symptomatic and disease modifying effects in a delayed start design. Delayed start design assumes similar symptomatic effects in patients treated early (0-36 Weeks) or later (>36 weeks) and there are no floor effects. Floor effects imply that the patients enrolled in the study have a certain degree of disease severity which will allow the drug to show the benefit. If the mean difference at the end of the study is significantly different between early and delayed start groups, then it would be concluded that the drug has disease modifying effects. Additional supportive evidence can also be gathered by comparing the slopes of placebo and treatment groups during 0-36 weeks. To ensure that the effects are persistent, it would be important to ensure that the rate of change in total UPDRS score is similar (parallelism) between early and delayed start groups (36-72 weeks).

Figure 4. Change from baseline in total UPDRS score for a drug with a mixture of symptomatic and disease modifying effects in a delayed start design study.



However, there are two scenarios, as shown in Figure 5, where primary analysis at the end of the active control phase would not be ideal for a drug with a mixture of symptomatic and disease modifying effects.

Figure 5. Scenarios where primary analysis at the end of the study would underestimate the disease modifying effect of a drug



Based on LSMEAN analysis at 72 weeks in ADAGIO, it is concluded that low dose (1 mg) has disease modifying effects while high dose (2 mg) does not (Figure 3). The sponsor and reviewer explored if the unexpected findings were due to issues with delayed start design as shown in Figure 5.

Figure 6 shows the change from baseline in total UDPRS score in patients with baseline total UPDRS scores of  $\leq 14$ , 14-19, 19-25.5 and  $> 25.5$ . There is no dose-response at 36 weeks in patients with baseline total UPDRS score greater than 25.5. These findings are similar to those observed in TEMPO study in which the mean baseline total UPDRS score was approximately 25 (Table 1).

Figure 6. Mean change from baseline in total UPDRS score by treatment group and baseline total UPDRS score quartiles in ADAGIO study.

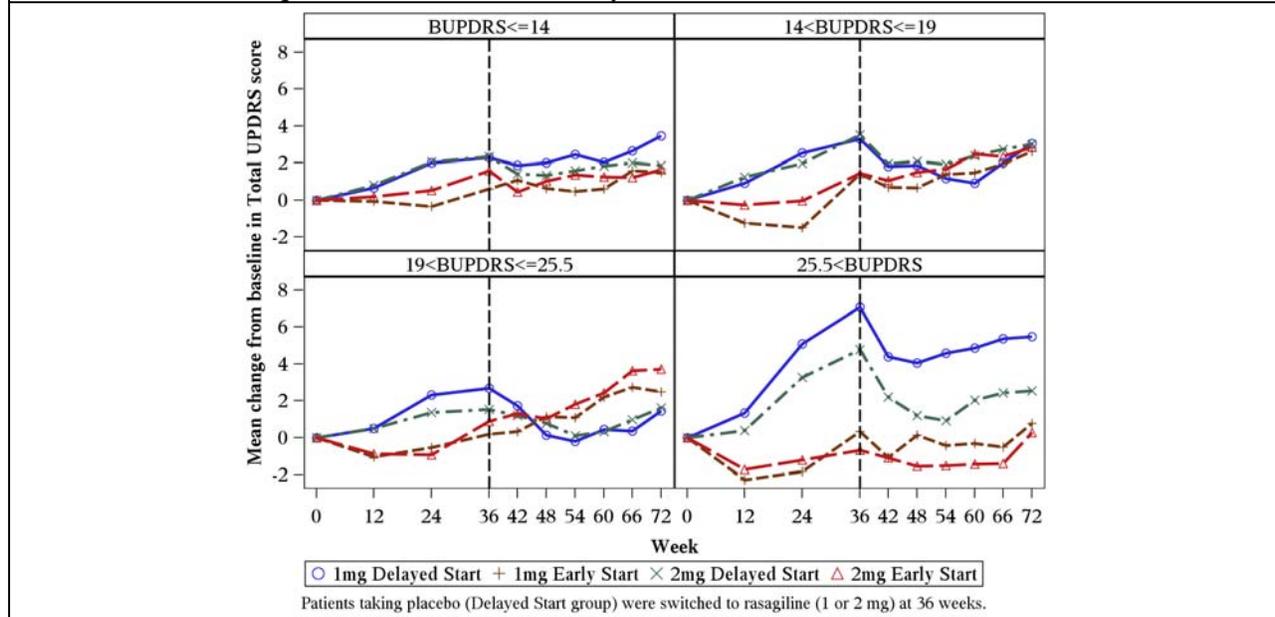


Figure 6 shows that

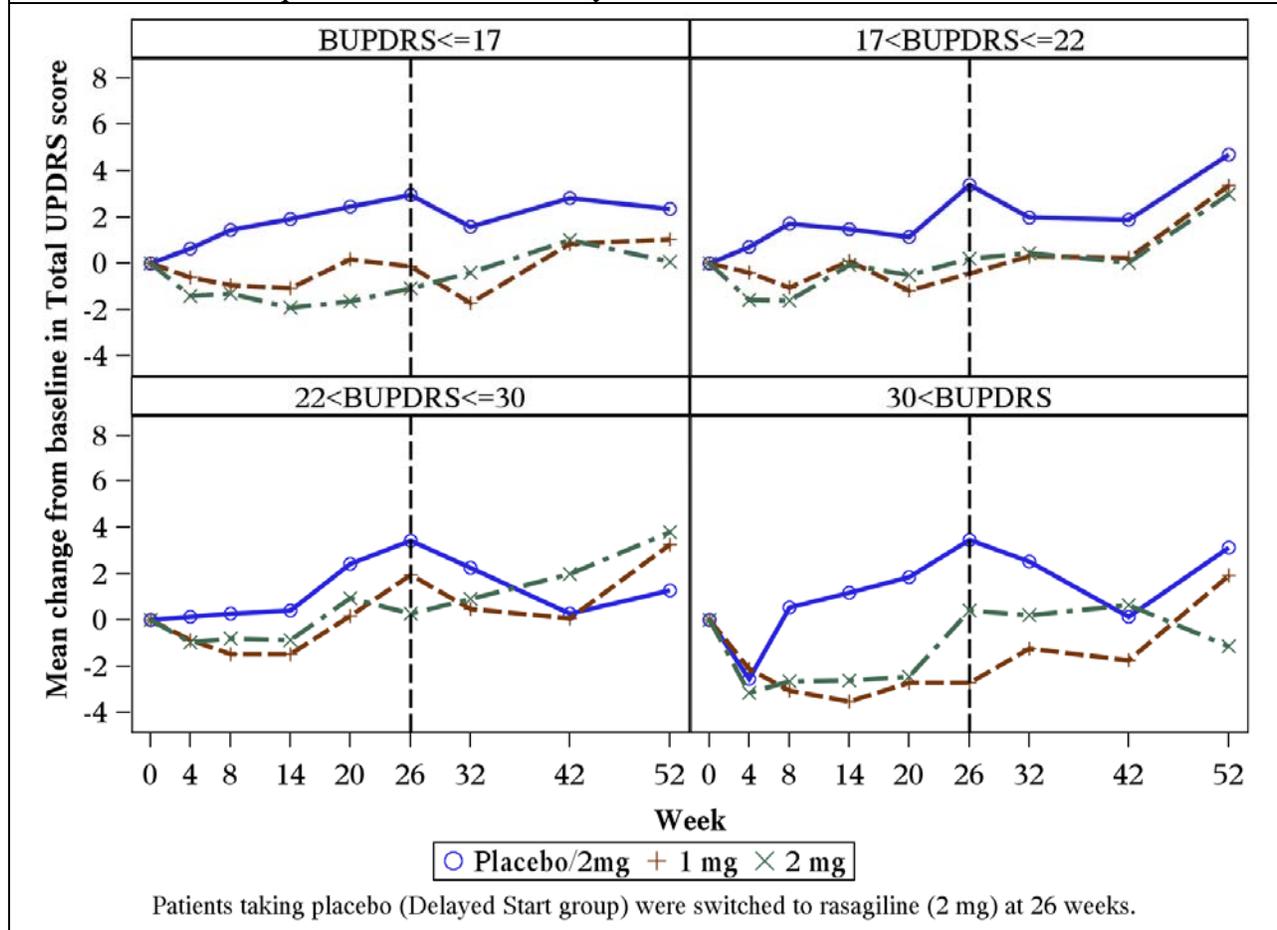
- The treatment effects of 1 mg are mainly observed in patients with baseline total UPDRS score greater than 25.5.
- The slope of disease progression in patients with baseline total UPDRS score greater than 25.5 is faster in comparison to other groups. The estimates of rate of progression in various groups, using linear mixed effects analysis, are shown in Table 3.

Table 3. Estimate of rate of change in total UPDRS score (slope per week) by treatment group and baseline total UPDRS quartiles.

	<b>Group</b>	<b>Baseline UPDRS Quartile</b>	<b>Slope, Per Week</b>
1mg Delayed Start		BUPDRS<=14	0.08091
		14<BUPDRS<=19	0.1473
		19<BUPDRS<=25.5	0.1255
		25.5<BUPDRS	0.2720
1mg Early Start		BUPDRS<=14	0.03587
		14<BUPDRS<=19	0.1150
		19<BUPDRS<=25.5	0.07891
		25.5<BUPDRS	0.1367
2mg Delayed Start		BUPDRS<=14	0.08350
		14<BUPDRS<=19	0.1042
		19<BUPDRS<=25.5	0.1119
		25.5<BUPDRS	0.2340
2mg Early Start		BUPDRS<=14	0.06048
		14<BUPDRS<=19	0.07315
		19<BUPDRS<=25.5	0.08112
		25.5<BUPDRS	0.05299

- Figure 7 shows the change from baseline in total UPDRS score by baseline total UPDRS quartiles in TEMPO study. There was no dose-response relationship between 1 and 2 mg in patients with baseline total UPDRS score greater than 30 in TEMPO study as shown in Figure 7. Also the slope of disease progression in various baseline total UPDRS quartiles in placebo group is not dependent on baseline total UPDRS score. These findings that the effects of influence of baseline total UPDRS score on treatment effect are not consistent in TEMPO and ADAGIO.

Figure 7. Mean change from baseline in total UPDRS score by treatment group and baseline total UPDRS score quartiles in TEMPO study.



**4.2 Based on the observation that the assumptions in delayed start design are not the reasons for lack of disease modifying effects of 2 mg, are there any other reasons for these findings?**

To understand the reasons for differences between 1 and 2 mg dose groups, the reviewer analyzed the baseline scores by treatment group and also longitudinal course of change from baseline total UPDRS score by country.

Table 4 shows that the baseline scores (UPDRS Motor, PIGD, Rigidity, Tremor, Total ADL, Bradykinesia) for 1 mg delayed start, 1 mg early start, 2 mg delayed start and 2 mg early start groups are comparable.

Table 4. Baseline scores in various treatment groups

Group	UPDRS Motor (Baseline)		UPDRS PIGD (Baseline)		UPDRS Rigidity (Baseline)		UPDRS Tremor (Baseline)		UPDRS Total (Baseline)		UPDRS ADL (Baseline)		UPDRS Bradykinesia (Baseline)	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
1mg Delayed Start	300	14.03	300	1.09	300	2.87	300	3.79	300	20.20	300	5.26	300	6.24
1mg Early Start	288	14.51	288	1.19	288	2.95	288	3.65	288	20.57	288	5.10	288	6.58
2mg Delayed Start	295	13.82	295	1.10	295	2.86	295	3.58	295	19.93	295	5.05	295	6.34
2mg Early Start	293	14.55	293	1.12	293	3.07	293	3.74	293	20.83	293	5.37	293	6.48

Figure 3 shows that the mean change from baseline in total UPDRS score in 2 mg delayed start group is lower when compared to 1 mg delayed start group during 0-36 weeks. Figure 8, Figure 9, Figure 10 and Figure 11 show the mean change from baseline in total UPDRS score by country. These graphs show that the change from baseline in total UPDRS score during 0-36 weeks is much slower in countries such as Israel, Netherlands, Spain.

Figure 8. Longitudinal change from baseline in total UPDRS score by country

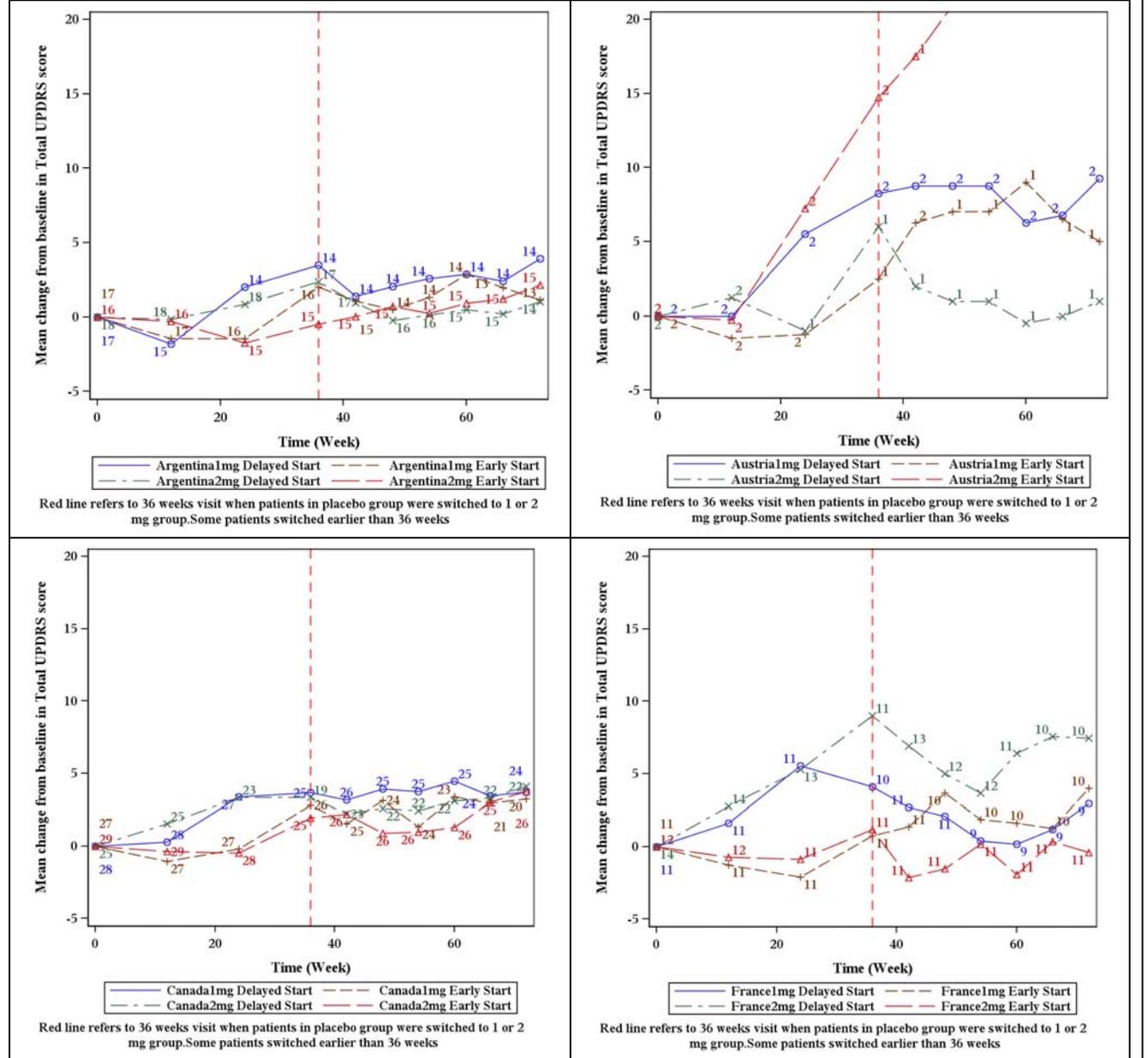


Figure 9. Longitudinal change from baseline in total UPDRS score by country

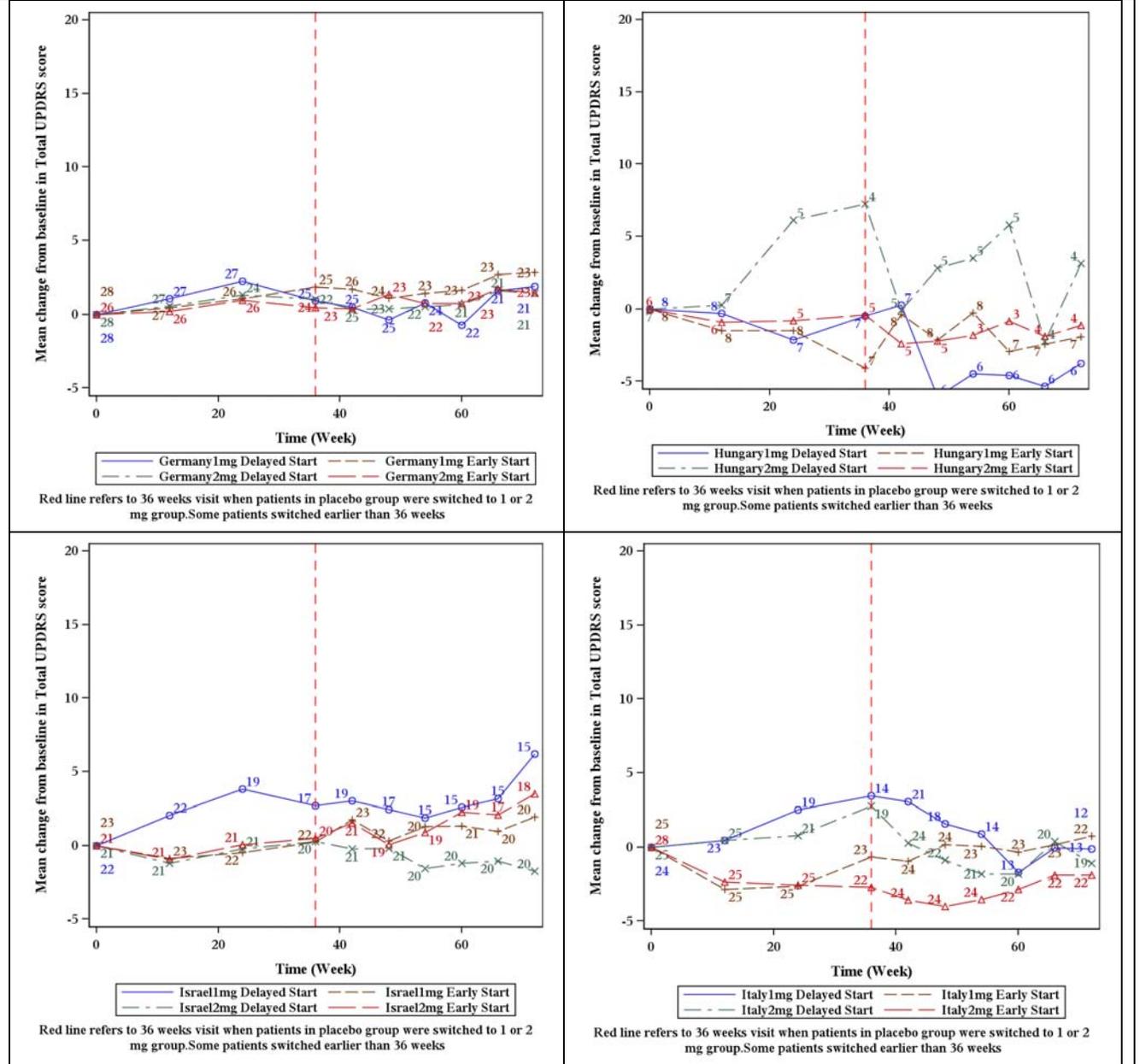


Figure 10. Longitudinal change from baseline in total UPDRS score by country

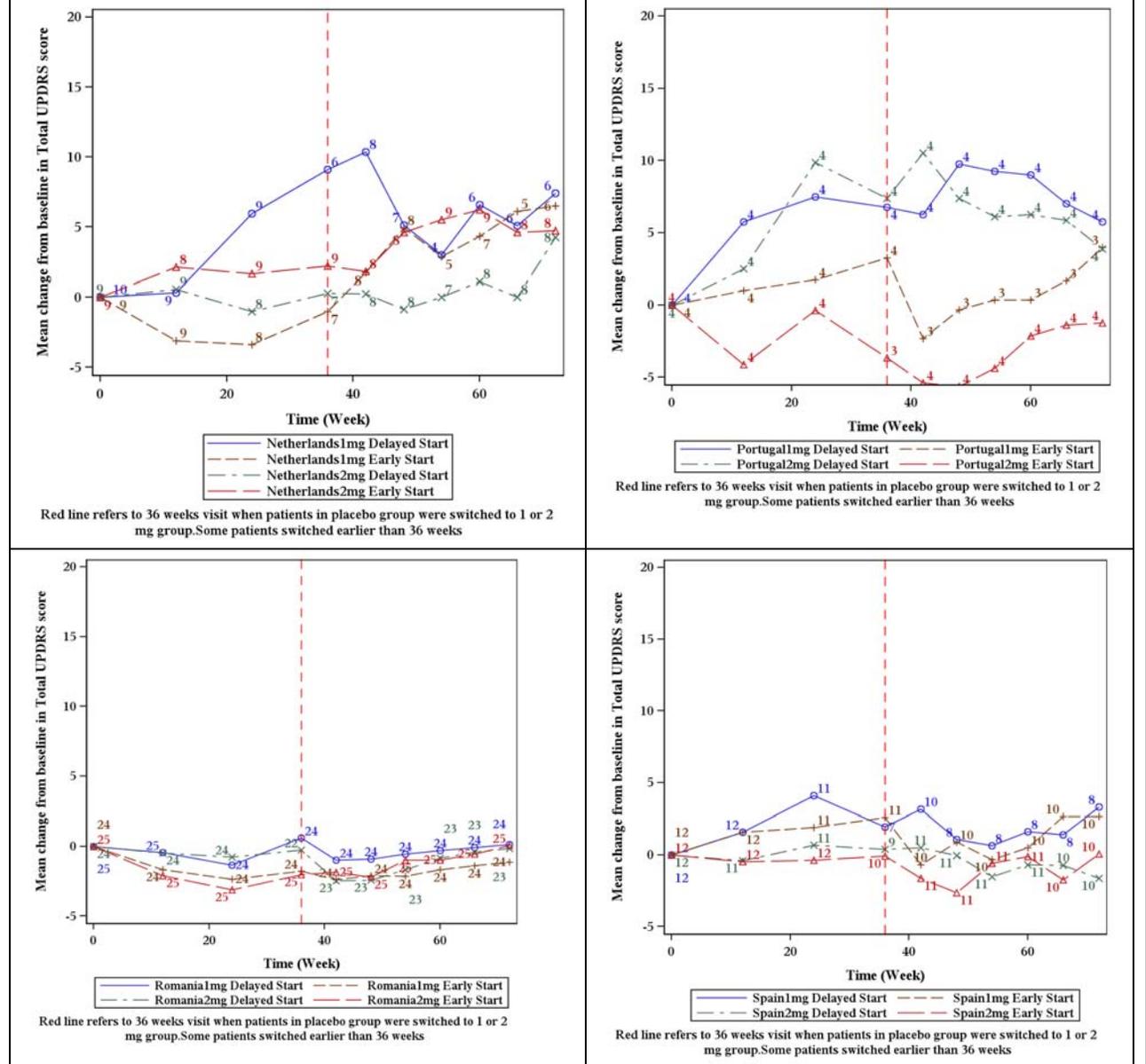


Figure 11. Longitudinal change from baseline in total UPDRS score by country

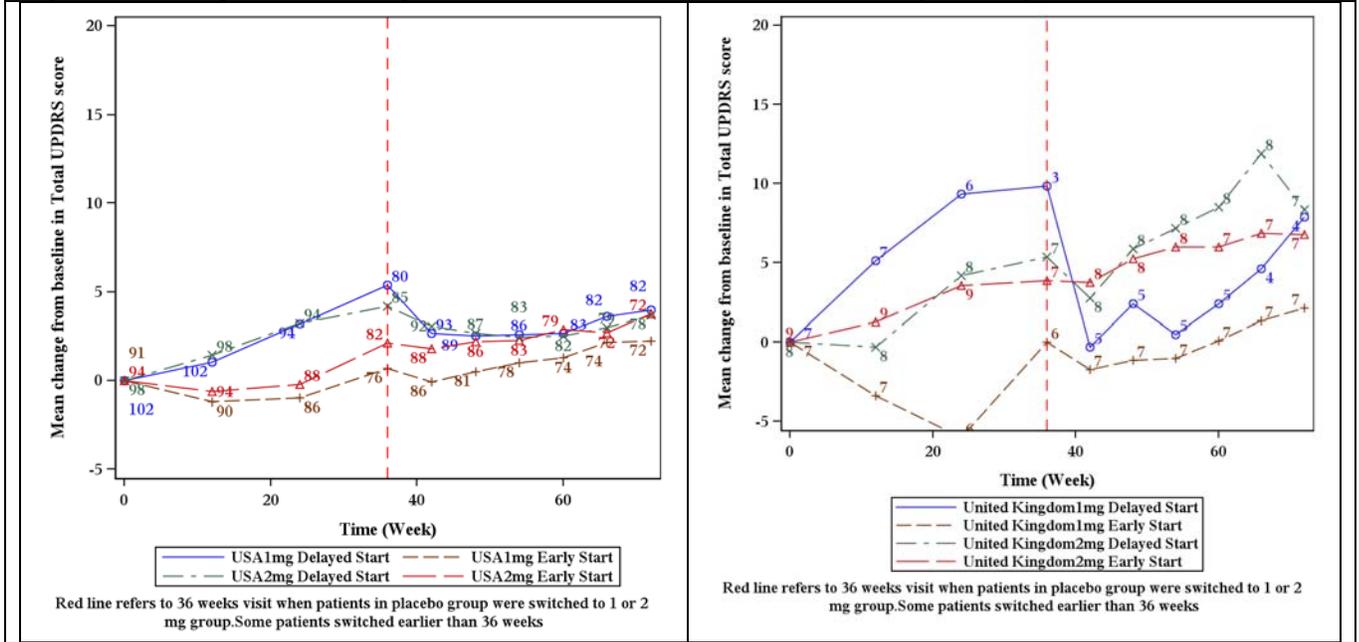
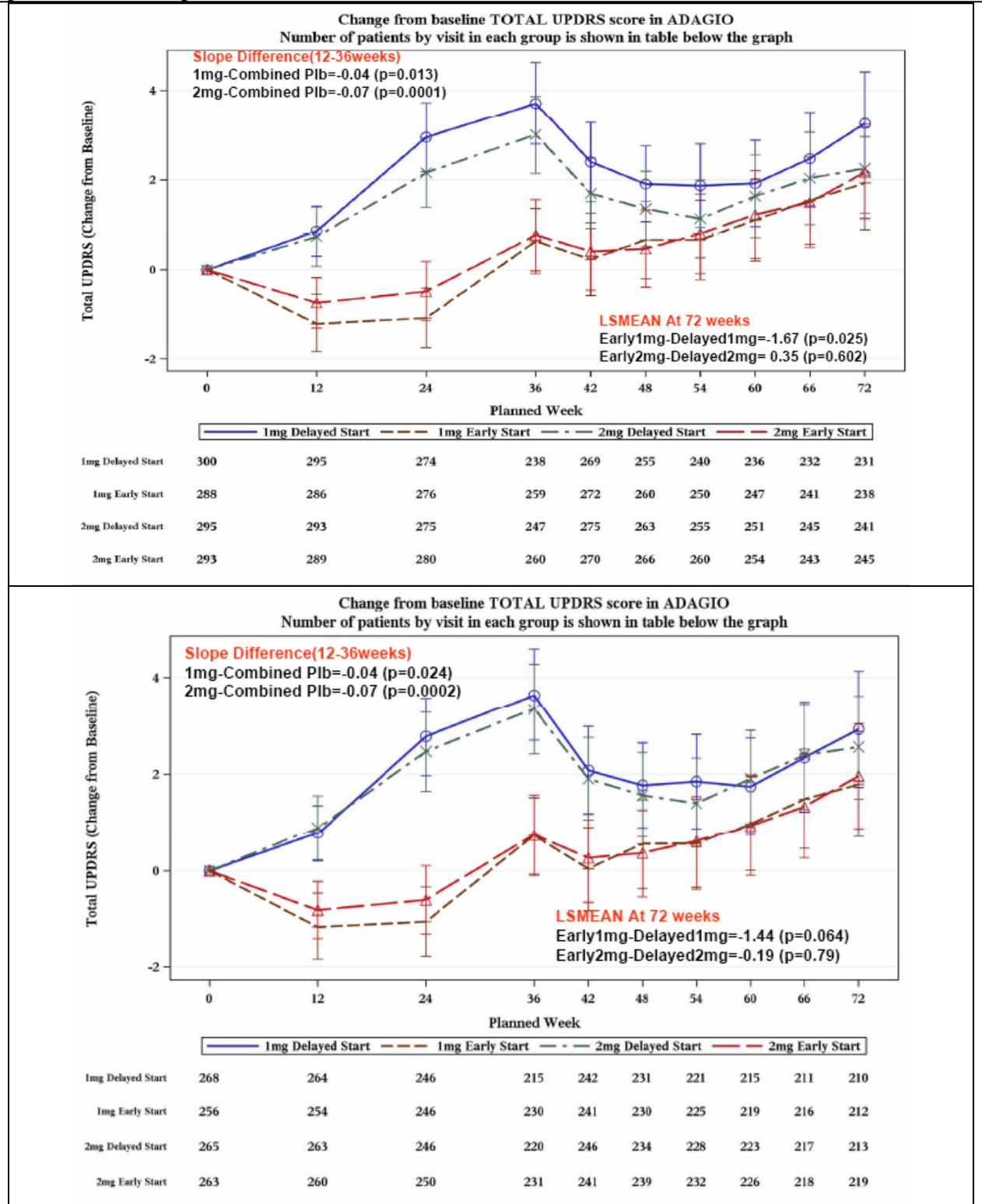


Figure 12 shows the longitudinal change from baseline total UPDRS score from all patients in the trial and after removal of data from Israel and Netherlands. In addition, the results of slope comparison between early and delayed start group (0-36 weeks) and mean difference between early and delayed start group at 72 weeks are shown in Figure 12. Exclusion of the data from these countries appears to influence the treatment effects of both 1 and 2 mg. The reasons for these findings are not clear.

Figure 12. Longitudinal change from baseline in total UPDRS score in (A) all patients (B) patients remaining after removal of data from Israel and Netherlands.







U.S. Department of Health and Human Services  
Food and Drug Administration  
Center for Drug Evaluation and Research  
Office of Translational Sciences  
Office of Biostatistics

# STATISTICAL REVIEW AND EVALUATION

## CLINICAL STUDIES

**NDA/BLA Serial Number:** 21641 (S0030)

**Drug Name:** Rasagiline (Azilect)

**Indication(s):** Parkinson's Disease: Slowing of Disease Progression

**Applicant:** Teva

**Date(s):** 12/23/2010

**Review Priority:** Priority

**Biometrics Division:** HFD-710

**Statistical Reviewer:** Tristan Massie, Ph.D.

**Concurring Reviewers:** Kun Jin, Ph.D., Team Leader  
Hsien Ming (James) Hung, Ph.D., Division Director

**Medical Division:** HFD-120 Division of Neurology

**Clinical Team:** Leonard Kapcala, M.D.  
Gerald Podskalny, M.D., Team Leader  
Russell Katz, M.D., Division Director

**Project Manager:** Stacy Metz

**Keywords:** Delayed Start; Multiple Imputation; Dose Response; Multiplicity

# Table of Contents

STATISTICAL REVIEW AND EVALUATION .....	1
LIST OF TABLES.....	3
LIST OF FIGURES.....	5
<b>1. EXECUTIVE SUMMARY .....</b>	<b>6</b>
<b>2. INTRODUCTION .....</b>	<b>9</b>
2.1 OVERVIEW.....	9
2.2 DATA SOURCES .....	10
<b>3. STATISTICAL EVALUATION .....</b>	<b>11</b>
3.2 EVALUATION OF EFFICACY .....	11
3.2.1 <i>TEMPO Study (Study 232)</i> .....	11
3.2.1.1 <i>Study Design and Statistical Methods</i> .....	11
3.2.1.2 <i>Patient Disposition</i> .....	16
3.2.1.3 <i>Sponsor’s Results for Placebo Controlled Phase</i> .....	19
3.2.1.4 <i>Active Phase Efficacy Results</i> .....	21
3.2.1.5 <i>Reviewer’s Assessment of the Impact of Missing Data on the Active Phase Results</i> .....	23
3.2.2 <i>ADAGIO (Study 500)</i> .....	26
3.2.2.1 Study Design and Analysis Plan .....	26
3.2.2.2 Subject Disposition .....	34
3.2.2.3 Baseline Demographic Characteristics.....	38
3.2.2.4 Sponsor’s Results.....	41
3.2.2.4.1 <i>Placebo Controlled Phase</i> .....	41
3.2.2.4.2 <i>ACTIVE PHASE</i> .....	44
3.2.2.4.2.1 <i>Hypothesis #2: Mean Difference at Week 72/Active Week 36</i> .....	44
3.3 EVALUATION OF SAFETY .....	63
<b>4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS .....</b>	<b>63</b>
4.1 GENDER, RACE, AGE, AND GEOGRAPHIC REGION .....	63
4.1.1 <i>Gender</i> .....	63
4.1.2 <i>Race</i> .....	67
4.1.3 <i>Age</i> .....	68
4.1.4 <i>U.S. vs. Non-U.S.</i> .....	71
4.2 OTHER SPECIAL/SUBGROUP POPULATIONS .....	74
4.2.1 <i>Baseline UPDRS Score</i> .....	74
<b>5. SUMMARY AND CONCLUSIONS .....</b>	<b>78</b>
5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE .....	78
5.2 CONCLUSIONS AND RECOMMENDATIONS .....	78
<b>I. APPENDIX 1: POST-HOC NATURAL HISTORY STAGGERED START ANALYSIS .....</b>	<b>79</b>
A. BACKGROUND AND SPONSOR’S RESULTS.....	79
B. REVIEWER’S COMMENTS ON NATURAL HISTORY STAGGERED START ANALYSIS .....	85
<b>II. APPENDIX 2: REFERENCES.....</b>	<b>88</b>

## LIST OF TABLES

Table 1 List of all Relevant Clinical Efficacy Trials .....	10
Table 2 TEMPO: Demographic Characteristics .....	16
Table 3 TEMPO: Termination Reasons by the Need for Additional* anti-PD Therapy .....	17
Table 4 TEMPO: Baseline Disease Characteristics .....	18
Table 5 TEMPO: Additional Baseline Disease Characteristics .....	18
Table 6 TEMPO: Need for Additional Anti-PD Therapy PC Phase.....	19
Table 7 Descriptive Statistics of Total UPDRS and Change from Baseline by Visit Using the Actual Visit Imputation Scheme .....	20
Table 8 TEMPO: Need for Additional Anti-PD Therapy Active Phase .....	21
Table 9 Efficacy Cohort Designation for Active Phase .....	21
Table 10 Mean Changes from Baseline in UPDRS Total (LOCF).....	22
Table 11 TEMPO: Mean Changes from Baseline in UPDRS Total (Observed Cases) .....	23
Table 12 TEMPO Active Phase Observed Mean Changes in UPDRS by Last Assessment Time .....	25
Table 13 Repeated Measures Analysis of Group Difference By Week in Active Phase .....	25
Table 14 Subject Disposition.....	35
Table 15 Summary of Discontinuations Prior to Active Phase and Early Transfers to Active Phase .....	35
Table 16 Eligibility for Primary Analysis of Active Phase .....	36
Table 17 Subject Disposition During Active Phase.....	36
Table 18 Reasons for Subject Premature Termination During Active Control Phase .....	37
Table 19 Reasons for Exclusion from ACTE Dataset .....	38
Table 20 Number of ITT Subjects Eligible for Active Phase Analysis .....	38
Table 21 Descriptive Statistics of Total UPDRS at Baseline -ITT Set.....	39
Table 22 Descriptive Statistics of Total UPDRS at Baseline -ACTE Set.....	39
Table 23 Potentially Important Race Difference between the Treatment Groups by Trial Phase .....	40
Table 24 Potentially Important Differences between the Treatment Groups in Baseline Variables by Trial Phase.....	40
Table 25 Descriptive statistics of Time from PD Diagnosis (ITT).....	41
Table 26 Model Estimates for Changes per Week (Slope) in Total UPDRS during PC Phase Efficacy ITT set .....	42
Table 27 Comparison of Changes per Week (Slope) for Rasagiline vs. Placebo during PC Phase (week 12-Week 36) Efficacy ITT Set .....	43
Table 28 Dependence of Week 72 Difference on Analysis Data Set Used .....	46
Table 29 Comparison of Changes from Baseline to Week 72 in Total UPDRS Score, 1 mg.....	49
Table 30 Baseline UPDRS score by Treatment Interaction: Baseline score coefficient estimates.....	49
Table 31 Comparison of Changes from Baseline to Week 72 in Total.....	52
Table 32 Group Differences in Changes from Baseline to Week 72 in Total UPDRS Scores – CO and PP Data Analysis Sets .....	52
Table 33 Multiple Imputation Sensitivity Analysis for Hypothesis 2 (Sponsor’s Analysis) .....	53
Table 34 Change from Baseline to Week 72 in Total UPDRS Score – .....	54
Table 35 Comparison of Changes from Baseline to Week 72 in Total.....	55
Table 36 Comparison of Mean UPDRS Change from Baseline between Active Phase Analysis Eligible and Ineligible Patients .....	56
Table 37 Change Per Week (Slope) in Total UPDRS Score during Active- .....	57
Table 38 Comparison of Changes per Week (Slopes) in Total UPDRS (Separate Datasets) .....	58
Table 39 Comparison of Group Differences in Changes per Week (Slopes) in Total UPDRS (Combined Dataset) .....	58
Table 40 ADAGIO: Week 36 Estimated Differences in UPDRS Change from Baseline by Gender.....	63
Table 41 ADAGIO: UPDRS change from baseline between 1mg groups by GENDER .....	64
Table 42 ADAGIO: 1MG Slope Differences Delayed-Early over Weeks 48 to 72 by Gender.....	65
Table 43 TEMPO: Week 52 Group Differences in UPDR Change by Gender .....	67
Table 44 ADAGIO: LS Mean differences in UPDRS Change at Week 36 by Age Group .....	68
Table 45 ADAGIO: Age Group specific Slope Differences over Weeks 48-72 for 1MG delayed-early .....	69
Table 46 TEMPO: Age Group Specific LS Group Mean Differences in UPDRS Change at Week 52 .....	71

Table 47 ADAGIO Distribution of ITT Subjects by Country .....	71
Table 48 Week 72 1MG Treatment Group Differences by Baseline UPDRS Quartile .....	76
Table 49 Week 72 2MG Treatment Group Differences by Baseline UPDRS Quartile .....	77
Table 50 TEMPO: Week 52 2MG Early vs. 2MG Delayed Group Differences by Baseline UPDRS Quartile .....	77
Table 51 TEMPO: Week 52 1MG Early vs. 2MG Delayed Group Differences by Baseline UPDRS Quartile .....	78
Table 52 ADAGIO: Natural History Staggered Start Disease Modification Estimates.....	83
Table 53 TEMPO: Natural History Staggered Start Disease Modification Estimates.....	84
Table 54 ADAGIO + TEMPO: Natural History Staggered Start Disease Modification Estimates .....	84
Table 55 ADAGIO: 1MG NHE Model 1 Analysis by Gender .....	86
Table 56 ADAGIO NHE Model 1 Residual Variance by Week and Baseline Quartile .....	87

## LIST OF FIGURES

Figure 1 TEMPO: Mean Change from Baseline in Total UPDRS (LOCF).....	22
Figure 2 ADAGIO Study Design .....	26
Figure 3 Sponsor’s Prespecified SAS Analysis Code for Hypothesis 2 .....	33
Figure 4 Change from Baseline in Total UPDRS Score during Placebo-Controlled.....	42
Figure 5 Mean UPDRS Change from Baseline Over Trial in 1 MG Groups (ACTE dataset).....	47
Figure 6 Mean UPDRS Change from Baseline Over Trial in 2 MG Groups (ACTE dataset).....	48
Figure 7 Change from Baseline in UPDRS during Active Phase (ACTE) for 1MG .....	61
Figure 8 Change from Baseline in UPDRS during Active Phase (ACTE) for 2MG .....	62
Figure 9 ADAGIO: Gender Specific Patterns of LS Mean UPDRS Changes by Visit for 1MG groups.....	65
Figure 10 ADAGIO: Gender Specific Patterns of LS Mean UPDRS Changes by Visit for 2MG groups.....	66
Figure 11 ADAGIO: Pattern of LS Means Changes from Baseline in UPDRS by Visit for 1MG by Age Group.....	69
Figure 12 ADAGIO Change from Baseline in UPDRS Profiles for 2MG by Age Group.....	70
Figure 13 ADAGIO: Change in UPDRS Mean Pattern US vs. Non-US 1MG Early and 1MG Delayed.....	73
Figure 14 ADAGIO Change from Baseline in UPDRS Total By Phase and Baseline UPDRS Total quartile.....	75
Figure 15 ADAGIO Change from Baseline in UPDRS Total By Phase in Highest Baseline UPDRS quartile .....	76

## 1. EXECUTIVE SUMMARY

The pre-specified primary analysis of the Change from baseline in UPDRS yielded an estimated difference at the end of the active phase, Week 72, of +1.42,  $p=0.0506$  for 1 mg delayed minus early and -0.179,  $p=0.8014$  for 2 mg delayed - early. If there is an effect on disease progression of 1mg Rasagiline the ADAGIO data suggests it may be inconsistent across sites, baseline UPDRS scores, and Gender. In particular, there were nominally significant separate interactions of the effects on change from baseline in UPDRS between Treatment group and each of the variables: Sites, baseline UPDRS score, and Gender. Thus, apparently, the overall 1 mg delayed-early difference is not a consistent or robust effect across many subgroups. The sponsor argued that because of the significance of the interactions with Site as well as baseline UPDRS score when the primary analysis was performed on the pre-specified combined dataset (all 4 groups analyzed at once) assumptions of the model were violated and, therefore, the separate datasets (1mg set: 1mg early and 1 mg delayed and 2 mg only set: 2mg early and 2 mg delayed) should be used instead. Restriction of the 1 mg analysis to the separate 1 mg dataset led to a p-value of 0.0250 for the comparison of means at Week 72, which is right at the required significance level, as adjusted to account for two dose comparisons.

Although, the overall 1 mg delayed –early difference appeared nominally significant when the analysis for that comparison excluded all 2 mg data there still appeared to be inconsistency of 1 mg treatment differences based on this separate dataset across the same subgroups (Site, Baseline UPDRS score, and Gender). Therefore, the post-hoc restriction to the separate dose-specific analysis datasets, although it just attains nominal significance, it does not address the inconsistency of the apparent 1 mg delayed-early effect across these important subgroups. This reviewer also found that one way of eliminating the significant Site by treatment group interaction was pooling sites by Country. The Country by Treatment group interaction was not significant ( $p=0.68$ ) and the 1 mg delayed-early difference at Week 72 based on the Countries adjusted model (instead of the more numerous Sites adjusted model) was +1.36,  $p=0.0873$  for the separate dataset ( $p=0.1178$  for the combined dataset).

Inconsistency of treatment group differences across sites was also suggested by the TEMPO data in the form of a statistically significant interaction between Treatment Groups and Sites.

A potential source of bias for the delayed start analysis is the randomized patients dropping out before reaching the active phase of the trial and, thereby, disturbing the remaining treatment groups' balance which was initiated by the randomization. In particular, the dropouts can leave the subset remaining for the analysis of the active phase with imbalances between the treatment groups in baseline demographics, and/or disease characteristics, and/or other measured or unmeasured variables. The presence of such imbalances means that any treatment group differences in outcome can likely no longer be attributed solely to the difference in treatment assignment. Actually, in the ADAGIO study although the sponsor calls the differences minor, the p-value for the 1mg Delayed vs. Early comparison of baseline UPDRS scores is 0.0563 based on ANOVA (Means: 19.1 Delayed vs. 20.5 Early or nonparametric Wilcoxon rank sum test  $p=.0523$ ) so the difference may be important (in the ITT set it was much bigger  $p=0.678$ ). Some

hypothesize that symptomatic effects may increase with increasing baseline score, in which case the observed group difference would create a bias in favor of the 1mg early group over the 1 mg delayed group. There were nominally significant baseline differences between 1 mg early and 1 mg delayed in the UPDRS Motor subscale, the UPDRS PIGD subscale and the UPDRS Brady subscale within the active phase dataset that were not significant within the ITT dataset. It should be acknowledged that some of these three scales have some items in common. There was also nominal statistical significance in the treatment group specific proportion of non-Caucasians although this was very small in both groups. These types of differences, particularly, in the UPDRS at baseline are exactly what one would not want to see for a delayed start trial because it means that the reported analysis of the active phase may be biased, in this case particularly for the 1 mg delayed vs. early comparison, upon which the trial seems to rest. A post-hoc analysis of treatment effect by baseline UPDRS quartiles suggests that when averaged over these quartiles (as a means of comparing treatment groups better balanced in terms of baseline UPDRS) the treatment group difference does not reach the appropriate significance level. The average 1mg treatment difference (delayed-early) over the four baseline UPDRS quartiles subgroups is 1.66 +/- .755 S.E.,  $p=0.0283 > 0.0250$  (based on the separate 1mg dataset). Even this analysis may not entirely correct the baseline imbalance.

The failure of the high dose to show an effect at the end of the active phase in the ADAGIO study forces one to consider the possibility of an uncommon U-shaped, or decreasing response with increasing dose, dose-response relationship. The 2 mg delayed group was numerically better than 2 mg early group (i.e., favoring delayed treatment) at week 72 in the overall Active Phase dataset (difference=.165,  $p=0.828$ ) and also in the U.S. subset (1.32,  $p=.278$ ). In fact, in most of the controlled clinical studies there isn't much evidence of an increasing response with dose, therefore, on that basis, one may consider pooling the two early rasagiline dose groups, 1 mg and 2 mg. This combined dose early rasagiline group sensitivity analysis also fails to show a significant effect compared to the combined delayed group at the end (Week 72) of the active phase, the estimated difference is -0.689 (early-delayed),  $p=0.1268$  (see Table 29). Suppose for the sake of argument that 2MG early group has the same effect as the nominally positive result observed for 1mg early group, -1.6799, based on the post-hoc separate dataset analysis, compared to delayed start. The probability that the maximum of two normally distributed variables (one for 1 mg early vs. delayed and one for 2 mg early vs. delayed) identically distributed with a mean of -1.6799 and standard deviation of .7469 (as observed for 1mg) is greater than .3565 (the observed 2mg early-delayed difference)  $= 1 - \Phi^2[(.3563 - -1.6799)/.7469] = .0064$ . Thus, it seems rather unlikely that the true 2mg delayed-early effect is the same as the observed nominally positive 1mg result based on the separate dataset analysis.

One possible explanation would be that the 1mg result is actually a type I error, i.e., a random high was observed by chance and there was no real delay of disease progression by 1 mg early. Also, considered on its own merits the 1 mg difference of 1.68 at week 72 seems to this reviewer rather modest compared to the 100+ point possible range of the UPDRS Total score and, furthermore, about half of the 1 mg delayed-early group difference at week 36 was lost by week 72 which suggests that it may all eventually disappear. Although the sponsor's test of nonlinearity had a p-value of  $0.089 > 0.05$  based on all four groups which led them to conclude the UPDRS change was linear over time in the Active Phase this reviewer found compelling

evidence of nonlinearity based on adding a quadratic term to test for nonlinearity with the 1 mg separate dataset. Such a quadratic test for nonlinearity was proposed in an early version of the protocol. A lack of linearity in the Active Phase undermines the intent of demonstrating parallelism of slopes, i.e., to show that separation of the delayed and early groups is maintained relatively constantly throughout the Active phase after Week 48 (by which time it was assumed any symptomatic effects would have been established). This nonlinearity would also seem to increase the uncertainty about whether the observed difference at Week 72 for 1 mg, such as it is, would persist beyond the scheduled end of the trial.

Significant nonlinearity was also found in the placebo controlled portion of the ADAGIO trial such that between Week 24 and 36 the 1 mg early slope was numerically bigger than the pooled placebo slope. Although over the entire placebo controlled phase the 1mg early slope was significantly better than pooled placebo slope that summary requires ignoring the nonlinearity which was manifested in the late apparent increase in the 1 mg early slope. Such a late increase in the 1 mg early slope relative to placebo doesn't fit the disease modification hypothesis.

There is no replication of dose specific results between the two trials. The TEMPO study had a 2mg early group and a 1 mg early group but only a 2mg delayed group. If one nevertheless compares 2 mg delayed to 1 mg early at week 52, which it seems the sponsor had every attention of doing when they drew up the analysis for the active phase, the result is clearly not nominally significant. If one accepts the analysis plan specified LOCF analysis of Change in UPDRS from baseline at Week 52 then 2mg early group appears nominally positive compared to 2 mg delayed but it's not clear that this can be considered to support the 1 mg results in ADAGIO when the 1mg early group in TEMPO did not differentiate from the delayed group. Not to mention the facts that the analysis of the active phase was originally designated as primarily for safety and exploratory for efficacy, there wasn't a clear single primary endpoint or primary analysis population for the active phase in TEMPO, and even the TEMPO 2 mg early group is not nominally significant compared to 2mg delayed in a standard MMRM analysis, which may be more appropriate than LOCF for reasons of potential bias caused by carrying forward data within a delayed start design, as well as the other standard reasons provided recently in the statistical literature.

There was only about 9% missing Week 72 data within the Active Phase subset (ACTE) of the ITT dataset, however, relative to the ITT dataset there was  $139/586 = 24\%$  for the 1MG groups. For the 2 mg groups there was  $41/507 = 8\%$  missing in the ACTE dataset and  $122/588 = 21\%$  in the ITT dataset. Fifteen percent (15%) of the ITT population was not eligible for ACTE and does not contribute any data to the two Active Phase primary analyses: the mean changes comparison at week 72 and the comparison of slopes during the Active Phase. The analysis of Completers' Changes from baseline in UPDRS Total at week 72 did not reach nominal significance for either 1 mg delayed vs. early (1.59,  $p=0.0374$ ;  $N=409$  Completers or 84% of the ACTE set) based on the separate dataset or for 2 mg delayed vs. early (0.088,  $p=0.8974$ ). Some prespecified multiple imputation sensitivity analyses achieved the nominal multiplicity adjusted level for the 1mg delayed-early comparison but not for the corresponding 2mg comparison. Other sensitivity analyses to missing data, such as imputing missing data with the delayed start mean regardless of the actual treatment group, did not achieve the required

significance level and, so, the overall picture for 1 mg is not a convincingly robust effect (to either missing data or other aspects).

If we consider that one of the two studies was primarily designed to assess symptomatic benefit and was analyzed for evidence of delaying disease progression, essentially as an afterthought, then at best we would be in the situation of one study plus confirmatory evidence. However, neither study is robustly positive and that doesn't even consider the complicating issue of the odd dose response pattern. Therefore, all things considered, there doesn't seem to be consistent compelling evidence of a delaying of Parkinson's disease progression provided by these two controlled Rasagiline trials.

## **2. INTRODUCTION**

### **2.1 Overview**

To date, the only available medications for Parkinson's disease (PD) are symptomatic, designed to ameliorate the clinical features of the illness. These therapies provide effective control of symptoms, particularly in the early phases of the disease. However, PD progresses over time, Parkinsonism worsens, the quality of drug response deteriorates, and the need for symptomatic medications increases. The majority of patients experience levodopa-related motor complications and new symptoms such as freezing, postural instability; falling and dementia that are not adequately controlled with existing medications. A major aim, therefore, is the limitation or halting of this process. A treatment which is able to slow down or stop neuronal death once PD has been diagnosed is the strategy currently receiving much attention from the scientific/medical community. Rasagiline is approved under the trade name Azilect, as a symptomatic anti-PD agent in the US, Canada, in the European Union and in an additional 13 countries. Rasagiline's efficacy as a symptomatic anti-PD agent was demonstrated in 3 pivotal trials. The first 6-month double-blind, randomized placebo-controlled phase of the TEMPO study demonstrated that rasagiline is efficacious as monotherapy in the treatment of early PD patients<sup>1</sup>. The 2 double-blind, randomized placebo - controlled adjunct therapy studies, PRESTO<sup>2</sup> and LARGO<sup>3</sup>, demonstrated the efficacy of rasagiline in the more advanced levodopa-treated PD population experiencing motor fluctuations. The IND number for this investigation of this drug is 45958.

There are only two clinical trials available for this drug that can be used to investigate the question of slowing of Parkinson's progression. They each incorporated a delayed start design: a placebo controlled period of 26 (TEMPO study) or 36 weeks (ADAGIO) followed by an active

treatment period of the same duration in which all patients were treated with the experimental treatment but the patients and investigators remained blinded in the active phase to the treatment assignment during the placebo controlled phase. The placebo controlled phase of TEMPO was used to support the indication of Rasagiline for symptomatic treatment of Parkinson’s disease. The analysis of the Active Phase of TEMPO was originally considered exploratory in terms of efficacy by the sponsor.

**Table 1 List of all Relevant Clinical Efficacy Trials**

Study	Phase and Design	Treatment Period	Follow-up Period	# of Subjects per Arm	Study Population
232 TEMPO	Phase 3 Delayed Start	26 weeks placebo controlled; 26 weeks active controlled (remain blinded to PC assignment)	52 weeks	1MG: 134/ 2MG: 132/ Placebo (2MG Delayed): 138	Mean Baseline UPDRS=25
500 ADAGIO	Phase 3 Delayed Start	36 weeks placebo controlled; 36 weeks active controlled (remain blinded to PC assignment)	72 weeks	1MG Delayed: 298 1MG Early: 288 2MG Delayed: 295 2MG Early: 293	Mean Baseline UPDRS=20.4

## 2.2 Data Sources

The primary endpoint data for TEMPO (Study 232) is located as follows.

[\\fdswa150\nonectd\n21641\N\\_000\2003-09-05\cr\datasets\TVP-1012-232A\UPDR\\_ORG.xpt](\\fdswa150\nonectd\n21641\N_000\2003-09-05\cr\datasets\TVP-1012-232A\UPDR_ORG.xpt)

The study report for TEMPO is located internally as follows.

<\\Cdsesub1\evsprod\NDA021641\0030\m5\53-clin-stud-rep\535-rep-effic-safety-stud\parkinsons\5351-stud-rep-contr\232\232-principal-text.pdf>

The primary endpoint data for ADAGIO (Study 500) is located internally as follows.

[\\Cdsesub1\evsprod\NDA021641\0008\m5\datasets\500\listings\upd\\_all.xpt](\\Cdsesub1\evsprod\NDA021641\0008\m5\datasets\500\listings\upd_all.xpt)

Study Report for ADAGIO is located internally as follows.

<\\Cdsesub1\evsprod\NDA021641\0030\m5\53-clin-stud-rep\535-rep-effic-safety-stud\parkinsons\5351-stud-rep-contr\500\tp-1012-500-report-body.pdf>

Programs related to analyses presented in the ISE such as the Natural History Estimate analysis are described in the following internal location.

<\\Cdsub1\evsprod\NDA021641\0030\m5\datasets\ise\analysis\programs\programs-documentation.pdf>

Only select files containing the sponsor's code for carrying out the analysis were submitted originally with the NDA. Others were obtained later after making specific requests for them.

### **3. STATISTICAL EVALUATION**

#### **3.2 Evaluation of Efficacy**

##### **3.2.1 TEMPO Study (Study 232)**

The study started on November 7, 1997 and the last patient completed the placebo phase on November 29, 1999. The last patient completed the active phase on July 21, 2000. The original protocol was dated June 9, 1997. There were 7 protocol amendments, the last of which was dated July 10, 2000. The Data analysis plan is dated 13 March 2001.

The primary objective of this study was to assess the safety and efficacy of rasagiline in PD subjects who are not receiving or requiring carbidopa/levodopa/ therapy. The primary efficacy measure was to be the change in total UPDRS score, calculated from baseline to 26 weeks, comparing rasagiline 1 mg/day and 2 mg/day with placebo.

##### **3.2.1.1 Study Design and Statistical Methods**

This was a multicenter, double-blind, placebo-controlled, parallel group, Phase III clinical trial for the efficacy, tolerability and safety of two doses of rasagiline mesylate in early Parkinson's disease (PO) subjects not treated with levodopa. Subjects were to be randomized to one of two (1 mg or 2 mg/day) dosages of rasagiline or placebo. There was to be a 1-week titration phase, followed by a 25 week maintenance phase and a 6-month active treatment extension.

Three hundred and sixty (360) early Parkinson's subjects not treated with levodopa were to be enrolled at approximately 27 sites. A minimum of 9 subjects were to be enrolled at each site.

#### **Sample Size Rationale**

The sole end-point used to assess the sample size required for this trial was the baseline to month six mean change in total UPDRS. Results of power calculations showed that a total of 120 patients enrolled in each of the 3 trial arms could provide adequate power to detect (at 5% significance level) a real difference between the changes of 3 total UPDRS points or more.

The power was estimated under the assumption that the pooled standard deviation of the change from baseline to the last visit of total UPDRS is between 7.40 (estimated from the lazabemide study - Annals of Neurology 1996) and 8.75 (estimated from the DATATOP study - PSG internal report). The statistical test used was the t-test comparing the 1 mg group to placebo and the 2 mg group to placebo using Hochberg's Step-up Bonferroni procedure for multiple comparisons, with an overall "experimentwise") two sided alpha level of 0.05. For a pooled standard deviation of 7.40 units, the estimated power was calculated to lie between 81% and 93% when the true effect of the 2 mg dose compared to placebo is 3 points and the true effect of the 1 mg dose compared to placebo is between 0 points and 3 points. With a pooled standard deviation of 8.75 units and under the same assumptions as to size of effect, the power was estimated to lie between 66% and 82% .

### **Blinded Sample Size Re-estimation**

To examine whether the variance estimate that was used in the above sample size sensitivity analysis was adequate, an assessment of its magnitude was to be performed after 1/3 and 1/2 of the patients completed 6 months of the double-blind phase. That assessment was to be done without breaking the blind. In the case that the upper bound (since the simple estimate will include the between groups variation) of the variance estimate was found to be much larger than the one projected, the sponsor reserved the right to upsize the study via protocol amendment.

### **Analysis of Placebo Controlled Phase**

The UPDRS assessment was to be administered during the PC Phase at Screening, Baseline weeks 4, 8, 14, 20, and 26, and during the Active Phase at weeks 32, 42, and 52.

Only data collected during the 6 month double-blind period was to be used for assessing the efficacy of rasagiline.

The primary efficacy endpoint for this study is the change in total UPDRS from baseline to the 6 month visit. Patients who require levodopa before the 6 month visit and any others who terminate prematurely from the study were to have their last observation carried forward. In the principal analysis, the baseline adjusted analysis of covariance was to be used for comparing the adjusted mean differences between the changes observed in each of the active drug groups versus placebo (two comparisons) incorporating terms for treatment and center. The treatment-by-center interaction term was to not be included in the model if it was not statistically significant (i.e., if  $p > .05$ ) In case of a significant treatment by center interaction data presentation was also to be done on a center by center basis.

## **STATISTICAL SIGNIFICANCE LEVEL**

The significance level for this study was to be 5% using two-tailed tests. The treatment effect of rasagiline was to be tested for significance by performing two comparisons for each endpoint: the group treated with 1 mg/day was to be compared to placebo and the group treated with 2 mg/day was also to be compared to placebo. Hochberg's Step-up Bonferroni method was to be used in order to maintain the experiment-wise type I error of 5% (two-tailed).

## **Active Treatment Period**

The protocol states that “At the end of the 6-months of double-blind treatment, all patients will be transferred to an active treatment phase for an additional 6 months. Exploratory data assessment will attempt to evaluate only the added long-term safety information.”

The active treatment period of the study begins at week 26. All remaining study drug from the double-blind phase was to be collected at this time. All subjects were to begin treatment with active drug. There was to be a one week titration for subjects originally assigned to placebo and sham titration for subjects originally assigned to one of the treatment groups. During the titration week, patients originally on either placebo or 1 mg per day of rasagiline were to receive 1 mg per day of rasagiline. Patients originally assigned to 2 mg per day of rasagiline were to receive 2 mg per day of rasagiline during this week. After this titration week, all patients were to begin treatment at the doses they were to be on for the remainder of the study (1 mg per day or 2 mg per day). The group assigned to 1 mg per day during the double-blind phase was to take 1 mg per day of rasagiline during the active treatment phase. Both the group originally assigned to 2 mg per day and the group originally assigned to placebo were to take 2 mg per day of rasagiline during the Active period. The drug supply and labeling was to remain blinded to prevent identification of subjects who had been assigned to placebo in the double-blind phase.

The Data Analysis Plan states that regarding the Active treatment phase, the following several cohorts can be identified for the purpose of data presentation and statistical analysis. Thus, there appears to be a lack of a clear “primary analysis” population for the active phase.

Active Extension Cohort (Active)- includes all the patients who have entered the active treatment phase.

Full DB PC Cohort -includes all patients who have entered the active treatment phase, after a duration of 6-months in the double-blind placebo-controlled phase.

Full 12 months Cohort- includes all the patients who completed the trial after 12 months (i.e., completed both the 6 months of DB PC portion and the 6 months of the active treatment phase).

Statistical tests were to use the changes from baseline to termination value (week 52) or last observed value before the onset of actual additional therapy, whichever came first. LOCF was to be used for patients that terminated the active phase before week 52 (and did not begin additional therapy).

Baseline adjusted analysis of covariance was to be used for comparing the adjusted mean differences between the changes observed in each of the active drug groups versus the original placebo (2 comparisons). The statistical model was to include, the effects of treatment and center and the baseline measurement as a covariate. The treatment-by-center interaction term was to not be included in the model if it was not statistically significant (i.e., if  $p > .05$ ).

Teva's Response dated May 12, 2011 to FDA Medical Reviewer's Questions.  
Do you agree that the protocol and Statistical Analysis Plan (SAP) for TEMPO did not specify a single primary efficacy endpoint?

**Teva's Response:**

“It is correct that no single primary efficacy endpoint was specified in the SAP for the Active Phase of TEMPO. “

Do you agree that the SAP was never submitted to the DNP for review prior to breaking the blind of the active treatment phase?

**Teva's Response:**

“The SAP was not submitted to the DNP prior to breaking the blind for the Active Phase. “

**EFFICACY ENDPOINTS ANALYSES**

Descriptive statistics and statistical significance tests, were to aim at detecting differences in disease progression, between each of the groups long-term treated with rasagiline (1 mg/day and 2mg/day) and the 'placebo-2mg' arm patients.

In order to explore the effect of rasagiline as anti-Parkinson monotherapy treatment, efficacy measurements taken after the onset of additional therapy, were not to be included in the statistical analyses, but were to be included in the data listings.

Efficacy measurements that were recorded during the active extension phase, include:

- UPDRS scales,
- Need for Levodopa (LD),
- Quality of Life (QOL),
- Clinical Global Impression (CGI),
- Timed Motor Tests.

UPDRS scales and the Need for Levodopa assessments were both conducted at each one of the active extension visits: week 32, week 42, week 52. The Quality of Life (QOL), Timed Motor Tests and the Clinical Global Impression (CGI) assessments were conducted during the active extension period at week 52.

Statistical Analysis of final follow-up visit (Week 58) data was to be performed on the FU cohort patients, defined previously.

Primary Active extension efficacy analyses were to include the UPDRS total score and the Need for Levodopa assessments. UPDRS sub scales, QOL, CGI and Timed motor Tests were regarded as secondary efficacy variables.

## **Pooling of Sites**

Note that due to a small number of patients in center no. 78 (Long Island Jewish Medical Center), this center was to be pooled with Mount Sinai Medical Center (center no. 65).

## **PRIMARY ANALYSES**

### **UPDRS score - Total**

The total UPDRS is the sum of scores of three sub-scales: Mentation (composed of 4 items), ADL (Activities of Daily Living - composed of 13 items, item 16 - tremor, which is composed of 2 items - right and left, will be averaged) and Motor (composed of 27 items). Overall, the total UPDRS is composed of 44 items, each item ranges from 0 to 4 points, hence the total UPDRS score ranges from 0 to 176 points. A higher UPDRS rating correspond to worse disease condition.

Missing items in the UPDRS scale were to be replaced according to the LOCF rule.

Efficacy evaluations were to use the changes from baseline (week 0), at each visit that was conducted before the onset of actual additional anti-PD treatment.

### **Statistical Tests**

#### **Last observed value Analysis**

Statistical tests were to use the changes from baseline to termination value (week 52) or last observed value before the onset of actual additional therapy, whichever came first. LOCF was to be used for patients that terminated the active phase before week 52 (and did not begin additional therapy).

Baseline adjusted analysis of covariance was to be used for comparing the adjusted mean differences between the changes observed in each of the active drug groups versus the original placebo (two comparisons). The statistical model was to include, the effects of treatment and center and the baseline measurement as a covariate. The treatment-by-center interaction term was to not be included in the model if it was not statistically significant (i.e. if  $p > 0.05$ ).

Analysis was to be performed on the Active extension cohort and Full 12 months cohort.

#### **Repeated Measures analysis**

Changes from baseline of total UPDRS, at each active extension phase's visit before the onset of additional therapy, were also to be analyzed using a repeated measures model.

The model was to include the treatment group, center, time in active phase and treatment by time interaction, as explanatory variables. Using -2 Log likelihood ratio test, the significance of the treatment group and its interaction with time, was to be tested. The SAS MIXED procedure was to be used to perform the analyses. The treatment by center interaction was to be first removed from the model, if not significant.

Analysis was to be performed on the Active extension cohort, FOB cohort and Full 12 months cohort.

#### **Follow up (FU) analysis**

Applying to the FU cohort patients, changes from baseline to follow-up visit, were to be analyzed. Baseline adjusted analysis of covariance was to be used for comparing the adjusted mean differences, of each of the original active drug groups versus the original placebo. The statistical model was to include, the effects of treatment and center

and the baseline measurement as a covariate. The treatment-by-center interaction was to not be included in the model if it was not statistically significant (i.e. if  $p > 0.05$ ).

### 3.2.1.2 Patient Disposition

Four hundred and seventy-three (473) patients were screened. Of these, 404 (84%) patients enrolled into this study in USA (28 centers) and in Canada (4 centers) and were randomly allocated to three treatment groups: 1 or 2 mg rasagiline or placebo. On average, mean disease duration in all treatment groups was one year at study entry: 0.94 year for the placebo, 0.93 year for the 1 mg rasagiline and 1.16 year for the 2 mg group (ranged from few days to 10.6 years).

**Table 2 TEMPO: Demographic Characteristics**

TVP-1012/232 Placebo-Controlled Phase		1 mg	2 mg	PLACEBO	All	
Height (cm)	All	N	134	132	138	404
		Mean	171.6	171.6	171.9	171.7
		Std	9.0	9.7	9.7	9.4
		Min	149.9	149.9	147.5	147.5
		Max	192.0	190.0	191.1	192.0
	Male	N	90	74	93	257
		Mean	176.2	178.1	176.7	176.9
		Std	6.2	6.9	7.3	6.9
		Min	165.0	155.0	154.2	154.2
		Max	192.0	190.0	191.1	192.0
	Female	N	44	58	45	147
		Mean	162.1	163.3	162.0	162.5
		Std	5.7	5.4	5.7	5.6
		Min	149.9	149.9	147.5	147.5
		Max	172.7	175.3	175.0	175.3
Weight (kg)	All	N	134	132	138	404
		Mean	77.6	80.7	76.8	78.3
		Std	14.0	14.9	14.8	14.6
		Min	46.4	50.9	45.9	45.9
		Max	121.4	140.0	131.8	140.0
	Male	N	90	74	93	257
		Mean	82.6	86.8	82.6	83.8
		Std	11.9	12.0	12.8	12.4
		Min	63.6	58.2	54.5	54.5
		Max	121.4	117.3	131.8	131.8
	Female	N	44	58	45	147
		Mean	67.2	72.9	65.0	68.8
		Std	12.2	14.7	11.3	13.4
		Min	46.4	50.9	45.9	45.9
		Max	95.5	140.0	93.1	140.0
Age (years)	N	134	132	138	404	
	Mean	61.6	60.4	60.5	60.8	
	Std	10.3	11.4	10.8	10.8	
	Min	33.0	32.0	38.0	32.0	
	Max	92.0	79.0	79.0	92.0	

Table 3 summarizes the termination reasons by treatment group and the need for LD therapy. One hundred and eleven (82.8%) patients on 1 mg rasagiline, 105 (79.5%) patients on 2 mg rasagiline and 112 (81.2%) patients on placebo completed the 6-month, placebo-controlled phase of the study without needing LD therapy. Patients, who failed to complete the placebo-controlled phase due to a need for LD and continued into the active-treatment phase, were not considered as early withdrawals. Termination reasons dichotomized by the need for LD are presented in Table 3. A total of 22 (5.4%) patients did not complete the initial 26 weeks of the study. Nine (6.7%), 8 (6.1%) and 5 (3.6%) patients on 1, 2 mg rasagiline and placebo, respectively, did not have a normal conclusion. The differences between treatment groups in the number of patients with premature termination or the time on study to termination were not statistically significant.

**Table 3 TEMPO: Termination Reasons by the Need for Additional\* anti-PD Therapy**

TVP-1012/232 Placebo-Controlled Phase		1 mg		2 mg		PLACEBO		All	
		N	%	N	%	N	%	N	%
<b>Need for Additional Therapy</b>	<b>Termination Reason</b>								
<b>No</b>	<b>Normal Completion</b>	111	93.3	105	95.5	112	97.4	328	95.3
	<b>Adverse Experience</b>	5	4.2	1	0.9	1	0.9	7	2.0
	<b>Failed to Return</b>	1	0.8	.	.	.	.	1	0.3
	<b>Subject Request</b>	2	1.7	2	1.8	2	1.7	6	1.7
	<b>Unsatisfactory Response</b>	.	.	1	0.9	.	.	1	0.3
	<b>Other</b>	.	.	1	0.9	.	.	1	0.3
	<b>All</b>		119	88.8	110	83.3	115	83.3	344
<b>Yes</b>	<b>Termination Reason</b>								
	<b>Normal Completion</b>	14	93.3	19	86.4	21	91.3	54	90.0
	<b>Adverse Experience</b>	.	.	1	4.5	.	.	1	1.7
	<b>Subject Request</b>	.	.	.	.	1	4.3	1	1.7
	<b>Unsatisfactory Response</b>	.	.	1	4.5	1	4.3	2	3.3
	<b>Protocol Violation</b>	.	.	1	4.5	.	.	1	1.7
	<b>Other</b>	1	6.7	.	.	.	.	1	1.7
<b>All</b>		15	11.2	22	16.7	23	16.7	60	14.9
<b>All</b>	<b>Termination Reason</b>								
	<b>Normal Completion</b>	125	93.3	124	93.9	133	96.4	382	94.6
	<b>AE</b>	5	3.7	2	1.5	1	0.7	8	2.0
	<b>Failed to Return</b>	1	0.7	.	.	.	.	1	0.2
	<b>Subject Request</b>	2	1.5	2	1.5	3	2.2	7	1.7
	<b>Unsatisfactory Response</b>	.	.	2	1.5	1	0.7	3	0.7
	<b>Protocol Violation</b>	.	.	1	0.8	.	.	1	0.2
	<b>Other</b>	1	0.7	1	0.8	.	.	2	0.5
<b>All</b>		134	100.0	132	100.0	138	100.0	404	100.0

\*assessed as a need for LD

Note: Copied from page 85 of study 232 study report

Baseline disease characteristics are displayed in Table 4 and Table 5. No statistically significant differences (ANOVA) were demonstrated between groups, except for UPDRS mental scale (p=0.0123) and a boundary significance of Severity of Illness scale (p=0.0508, for baseline Severity of Illness see Table 5).

*Reviewer's Comment: Note that the cited differences are largest between the two groups most relevant during the Active Phase (2mg early start and placebo/2mg delayed start).*

**Table 4 TEMPO: Baseline Disease Characteristics**

TVP-1012/232 Placebo-controlled Phase		1 mg	2 mg	PLACEBO	All
<b>Total UPDRS</b>	<b>N</b>	134	132	138	404
	<b>Mean</b>	24.69	25.89	24.54	25.03
	<b>Std</b>	11.25	9.54	11.61	10.84
	<b>Min</b>	5.50	10.50	5.50	5.50
	<b>Max</b>	75.00	53.50	61.00	75.00
<b>UPDRS Mental 1-4</b>	<b>N</b>	134	132	138	404
	<b>Mean</b>	0.94	1.20	0.79	0.98
	<b>Std</b>	1.11	1.27	1.08	1.16
	<b>Min</b>	0.00	0.00	0.00	0.00
	<b>Max</b>	4.00	6.00	5.00	6.00
<b>UPDRS ADL 5-17</b>	<b>N</b>	134	132	138	404
	<b>Mean</b>	5.90	6.73	6.16	6.26
	<b>Std</b>	3.35	3.22	3.53	3.38
	<b>Min</b>	0.50	0.50	0.50	0.50
	<b>Max</b>	17.00	19.50	20.00	20.00
<b>UPDRS Motor 18-44</b>	<b>N</b>	134	132	138	404
	<b>Mean</b>	17.85	17.95	17.59	17.80
	<b>Std</b>	8.89	7.52	8.84	8.43
	<b>Min</b>	4.00	4.00	3.00	3.00
	<b>Max</b>	58.50	36.50	46.00	58.50

Note: This table was copied from sponsor's study report, page 107

**Table 5 TEMPO: Additional Baseline Disease Characteristics**

TVP-1012/232 Placebo-Controlled Phase		1 mg	2 mg	PLACEBO	All
<b>Severity of Illness (Baseline)</b>	<b>N</b>	133	132	136	401
	<b>Mean</b>	1.66	1.83	1.65	1.71
	<b>Std</b>	0.68	0.61	0.67	0.66
	<b>Min</b>	0.00	0.00	0.00	0.00
	<b>Max</b>	3.00	3.00	3.00	3.00

Note: Copied from study report page 108

There were no significant differences between 2mg delayed and 2 mg early groups in the Active Phase Analysis Data Set that were not present in the ITT dataset, except for a difference in a metabolic / endocrine information question at baseline.

**Table 6 TEMPO: Need for Additional Anti-PD Therapy PC Phase**

TVP-1012/232 Active-Treatment Phase	Treatment Group						All	
	1 mg		2 mg		Placebo/2 mg		N	%
	N	%	N	%	N	%		
<b>All</b>	<b>124</b>	<b>100.0</b>	<b>124</b>	<b>100.0</b>	<b>132</b>	<b>100.0</b>	<b>380</b>	<b>100.0</b>
<b>Did not Need Additional Therapy in PC* Phase</b>	<b>110</b>	<b>88.7</b>	<b>105</b>	<b>84.7</b>	<b>111</b>	<b>84.1</b>	<b>326</b>	<b>85.8</b>
<b>Needed Additional Therapy in PC Phase</b>	<b>14</b>	<b>11.3</b>	<b>19</b>	<b>15.3</b>	<b>21</b>	<b>15.9</b>	<b>54</b>	<b>14.2</b>

\*Placebo-Controlled

### 3.2.1.3 Sponsor's Results for Placebo Controlled Phase

The mean total UPDRS scores at baseline for all randomized patients were similar across all treatment groups (24.7, 25.9 and 24.5 for the 1 mg, 2 mg and placebo groups respectively,  $p=0.5385$ ). Following 26 weeks of treatment, the change from baseline UPDRS differed significantly between either of the active-treatment group and the placebo ( $p<0.0001$  for both contrasts using Hochberg's Step-up Bonferroni procedure for multiple comparisons). The adjusted mean change from baseline in total UPDRS score was -0.13 (95% CI:[-1.16, 0.91]) for the 1 mg group and 0.51 (95% CI:[-0.55, 1.57]) for the 2 mg group. Patients receiving placebo showed an increase of 4.07 (95% CI:[3.04, 5.10]) points. Thus, the treatment effect exerted by 1 and 2 mg rasagiline was -4.20 (95% CI:[-5.66,-2.73]) and -3.56 (95% CI:[-5.04,-2.08]), respectively.

**Table 7 Descriptive Statistics of Total UPDRS and Change from Baseline by Visit Using the Actual Visit Imputation Scheme**

TVP-1012/232 Placebo-Controlled Phase		Total UPDRS				Total UPDRS (Change from Baseline)			
		1 mg	2 mg	PLACEBO	All	1 mg	2 mg	PLACEBO	All
Baseline	N	134	132	138	404	134	132	138	404
	Mean	24.69	25.89	24.54	25.03	0.00	0.00	0.00	0.00
	Std	11.25	9.54	11.61	10.84	0.00	0.00	0.00	0.00
	Min	5.50	10.50	5.50	5.50	0.00	0.00	0.00	0.00
	Max	75.00	53.50	61.00	75.00	0.00	0.00	0.00	0.00
Week 4	N	134	132	137	403	134	132	137	403
	Mean	23.68	24.47	24.41	24.19	-1.01	-1.41	-0.24	-0.88
	Std	11.59	10.31	12.11	11.35	4.57	4.34	5.49	4.85
	Min	3.50	3.50	5.50	3.50	-15.00	-11.50	-17.50	-17.50
	Max	75.50	53.00	83.00	83.00	11.00	9.50	22.00	22.00
Week 8	N	128	127	136	391	128	127	136	391
	Mean	22.96	24.13	25.09	24.08	-1.59	-1.49	0.96	-0.67
	Std	11.62	10.74	12.44	11.64	5.31	5.01	6.13	5.63
	Min	3.50	3.50	5.00	3.50	-31.00	-18.00	-12.50	-31.00
	Max	62.50	61.00	70.00	70.00	11.50	14.00	25.50	25.50
Week 14	N	125	122	135	382	125	122	135	382
	Mean	22.99	24.07	25.39	24.18	-1.56	-1.30	1.26	-0.48
	Std	11.40	10.49	12.69	11.61	5.77	5.08	6.26	5.87
	Min	3.50	1.00	5.00	1.00	-27.50	-11.50	-14.50	-27.50
	Max	61.50	56.00	64.50	64.50	12.00	12.00	21.00	21.00
Week 20	N	118	117	120	355	118	117	120	355
	Mean	23.46	24.41	25.20	24.36	-0.85	-0.76	1.93	0.12
	Std	12.11	11.05	12.01	11.72	6.45	5.53	6.67	6.36
	Min	3.50	3.00	6.50	3.00	-37.00	-14.50	-19.00	-37.00
	Max	75.00	58.00	66.50	75.00	15.00	16.00	24.00	24.00
Week 26/Termination	N	115	107	115	337	115	107	115	337
	Mean	24.09	24.57	26.24	24.97	-0.00	-0.05	3.31	1.11
	Std	12.21	11.52	12.71	12.17	7.09	5.56	7.44	6.93
	Min	4.00	3.50	5.00	3.50	-39.00	-14.00	-18.50	-39.00
	Max	60.00	58.00	65.00	65.00	26.00	21.00	23.50	26.00

Note: This table was copied from page 96 of sponsor's study report

### 3.2.1.4 Active Phase Efficacy Results

Of the 380 patients who entered the active-treatment phase (active-treatment cohort), nine patients who received additional dopaminergic therapy or withdrew immediately following entrance to active-treatment phase (before the first efficacy assessment) were not included in the efficacy analysis. The other 371 patients (92%) were included in the efficacy analysis (efficacy cohort) of the active treatment phase.

**Table 8 TEMPO: Need for Additional Anti-PD Therapy Active Phase**

TVP-1012/232 Active-treatment Phase	1 mg		2 mg		Placebo/2 mg		All	
	N	%	N	%	N	%	N	%
<b>All</b>	<b>124</b>	<b>100.0</b>	<b>124</b>	<b>100.0</b>	<b>132</b>	<b>100.0</b>	<b>380</b>	<b>100.0</b>
<b>Started Additional Therapy</b>								
<b>No</b>	90	72.6	84	67.7	95	72.0	269	70.8
<b>Yes</b>	34	27.4	40	32.3	37	28.0	111	29.2

Note: This table was copied from page 167 of sponsor's study report

**Table 9 Efficacy Cohort Designation for Active Phase**

TVP-1012/232 Active-Treatment Phase	1 mg		2 mg		Placebo/2 mg		All	
	N	%	N	%	N	%	N	%
<b>All</b>	<b>124</b>	<b>100.0</b>	<b>124</b>	<b>100.0</b>	<b>132</b>	<b>100.0</b>	<b>380</b>	<b>100.0</b>
<b>Efficacy Cohort</b>								
<b>No</b>	2	1.6	5	4.0	2	1.5	9	2.4
<b>Yes</b>	122	98.4	119	96.0	130	98.5	371	97.6

Note: This table was copied from page 169 of sponsor's study report

The 52-week mean ( $\pm$ Std) changes from baseline were 3.01 (8.26), 1.97 (7.49) and 4.17 (8.83) for the 1 mg, 2 mg and placebo/2 mg treatment groups, respectively (Table 10). The difference between each of the long-term rasagiline treatment groups (1 and 2 mg) and the placebo/2 mg group (two contrasts) was statistically significant ( $p=0.046$  and  $p=0.024$ , respectively). The LOCF imputation was applied to account for missing data, early discontinuation and for measurements taken after the initiation of additional anti-PD therapy. Due to the influence of outlier patients (especially patient #198 in the 1 mg group), non-parametric testing was used post-hoc. The sponsor called patient id's 198, 344, and 388 outliers and also investigated excluding center 35. This reviewer found  $p=.06$  for the 2mg early vs. delayed comparison without PAT ID 344, using site effects as specified in the SAP, (other outliers were in 1MG which was left out of this analysis). The median changes from baseline were 3, 1.5 and 3.5 for the 1, 2 and placebo/2 mg treatment groups, respectively. The sponsor concludes that the difference gained between the placebo and 2 mg group at the end of the placebo-controlled phase (Week 26) was sustained for additional 26 weeks although both groups were treated with 2 mg rasagiline during that period. This difference was nominally significant (Figure 1,  $p=0.024$ ).

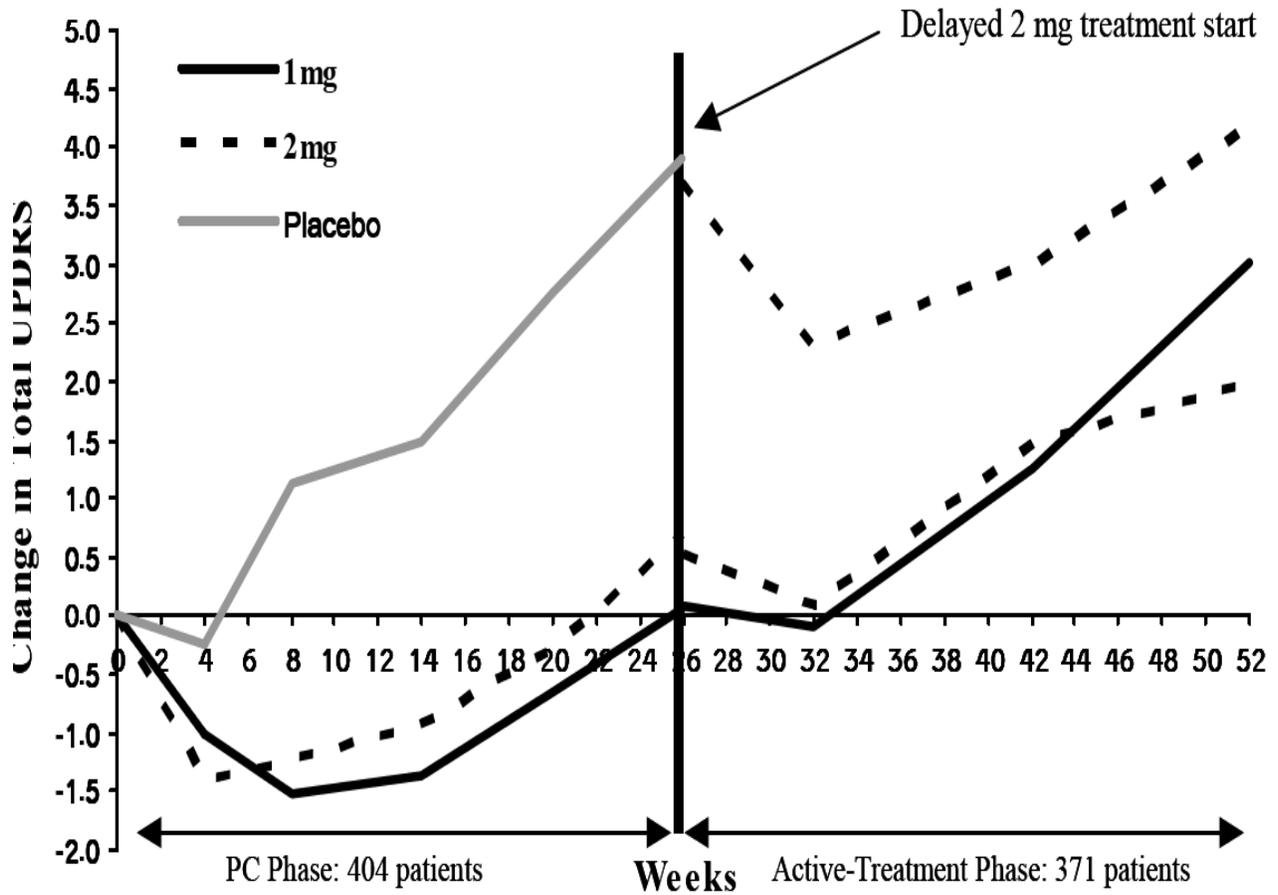
Table 10 summarizes the mean change from baseline in UPDRS total score at the baseline, intermediate, and final visit using last observation carried forward when the relevant post-baseline visit was missing.

**Table 10 Mean Changes from Baseline in UPDRS Total (LOCF)**

TVP-1012/232 Tempo		Total UPDRS				Total UPDRS (Change from Baseline)				Total UPDRS (Change from Week 26)			
		1 mg	2 mg	Placebo/2mg	All	1 mg	2 mg	Placebo/2mg	All	1 mg	2 mg	Placebo/2mg	All
Week 0 (Baseline)	N	124	124	132	380								
	Mean	24.52	25.44	23.77	24.56								
	Median	22.25	24.00	21.75	23.00								
Week 26/Placebo-Controlled Phase Termination	N	124	122	132	378	124	122	132	378				
	Mean	24.50	25.80	27.59	26.00	-0.03	0.50	3.82	1.49				
	Median	22.50	24.00	24.75	24.00	0.25	0.00	3.00	1.00				
Week 52/Last visit before Additional Therapy	N	122	119	130	371	122	119	130	371	122	119	130	371
	Mean	27.45	27.10	28.02	27.54	3.01	1.97	4.17	3.08	2.92	1.47	0.45	1.59
	Median	27.50	26.50	24.25	26.50	3.00	1.50	3.50	2.50	2.00	1.00	0.00	1.50

Note: This table was copied from page 173 of sponsor’s study report

**Figure 1 TEMPO: Mean Change from Baseline in Total UPDRS (LOCF)**



Note: Figure copied from page 174 of sponsor’s study report.

In addition, descriptive statistics of total UPDRS scores by actual weeks in the active-treatment phase (As Is, no LOCF) are presented in Table 11. This reviewer notes that the 2mg early – 2mg delayed group difference in total UPDRS change from baseline at week 52 is considerably smaller for the Observed Cases than in the LOCF imputed analysis. The LOCF analysis may be biased since the early group had the advantage of early treatment and earlier assessments would be more likely to show this than later ones (OC: 2.77 vs. 1.55 ; LOCF: 4.17 vs. 1.97). This pattern is true of the group medians as well.

**Table 11 TEMPO: Mean Changes from Baseline in UPDRS Total (Observed Cases)**

TVP-1012/232 Active-treatment Phase		Total UPDRS				Total UPDRS (Change from Baseline)				Total UPDRS (Change From Week 26)			
		1 mg	2 mg	Placebo/2 mg	All	1 mg	2 mg	Placebo/2 mg	All	1 mg	2 mg	Placebo/2 mg	All
Week 26/ PC Phase Termination	N	122	119	130	371	122	119	130	371	0	0	0	0
	Mean	24.54	25.63	27.56	25.95	0.09	0.51	3.71	1.49	.	.	.	.
	Median	22.50	24.00	24.75	24.00	0.50	0.00	2.75	1.00	.	.	.	.
	Std	12.30	11.48	13.33	12.45	6.58	5.87	7.36	6.83	.	.	.	.
	Min	4.00	3.50	5.00	3.50	-39.00	-14.00	-18.50	-39.00	.	.	.	.
	Max	60.00	58.00	65.00	65.00	13.50	21.00	23.50	23.50	.	.	.	.
Week 32	N	122	117	129	368	122	117	129	368	122	117	129	368
	Mean	24.36	25.12	26.20	25.25	-0.09	0.06	2.35	0.81	-0.18	-0.41	-1.41	-0.68
	Median	23.25	23.00	23.00	23.00	-0.50	0.00	1.50	0.50	-0.50	0.00	-1.00	-0.50
	Std	13.67	11.36	13.62	12.95	7.76	6.59	7.86	7.51	5.07	5.02	5.32	5.16
	Min	2.50	3.00	3.50	2.50	-44.50	-12.50	-25.00	-44.50	-15.00	-19.50	-18.50	-19.50
	Max	78.00	59.00	64.50	78.00	20.50	30.50	28.00	30.50	18.50	14.00	13.00	18.50
Week 42	N	102	94	102	298	102	94	102	298	102	94	102	298
	Mean	24.25	25.06	22.87	24.03	0.49	1.35	1.52	1.11	0.87	1.48	-1.19	0.36
	Median	22.50	23.25	20.75	22.00	0.00	0.75	2.00	1.00	1.00	1.00	-1.00	0.00
	Std	13.32	11.74	11.47	12.21	6.73	7.18	7.68	7.20	5.24	4.66	5.81	5.38
	Min	3.50	6.00	6.00	3.50	-30.00	-12.50	-18.00	-30.00	-16.00	-8.00	-17.50	-17.50
	Max	80.50	70.00	69.00	80.50	17.50	43.50	38.00	43.50	20.50	22.50	16.00	22.50
Week 52	N	89	82	91	262	89	82	91	262	89	82	91	262
	Mean	25.26	24.89	23.70	24.60	2.24	1.55	2.77	2.21	2.87	1.94	0.10	1.61
	Median	24.50	25.50	21.00	23.00	2.50	1.25	2.00	2.00	2.00	2.00	-0.50	1.50
	Std	14.22	11.05	12.40	12.63	8.76	6.86	9.44	8.45	6.42	5.20	7.45	6.54
	Min	2.50	3.50	3.50	2.50	-37.00	-10.00	-19.00	-37.00	-15.00	-10.50	-23.50	-23.50
	Max	70.00	58.50	61.00	70.00	27.50	22.50	45.00	45.00	31.00	15.00	24.00	31.00

Note: This Table copied from page 507 of sponsor's study report

### 3.2.1.5 Reviewer's Assessment of the Impact of Missing Data on the Active Phase Results

In the Completers population, defined as those who had a UPDRS assessment at week 52, the estimated mean UPDRS Change from baseline was 2.54 (+/- 0.8 S.E.) for PLACEBO (N=91) and 0.79 (+/- 0.82 S.E.) for 2 MG (N=82), thus, the estimated treatment difference on the UPDRS change at Week 52 was -1.75 (+/- 1.11 S.E.) favoring 2MG numerically, p=0.1153 . The 1MG

vs. Placebo/2MG difference was  $-1.10 \pm 1/07$  (S.E.),  $p=.3057$ , which numerically but not significantly favored 1MG early.

Table 12 displays Mean Changes in UPDRS by Last UPDRS Assessment Time. The pattern suggests that the LOCF analysis may be biased because the group difference is smaller in completers than for the LOCF analysis and also there is a large difference in those last assessed at week 32 which would be carried forward in the analysis, despite the fact that any symptomatic effect in the placebo group/2mg is probably not fully established this early (week 6) in the active phase portion. A repeated measures analysis also does not find a significant difference at week 52 for 2MG early vs. placebo/2MG delayed. The difference was  $2.05 \pm 1.04$  S.E.,  $p=.0501$  when adjusted for site as specified in the SAP. This was a common repeated measures analysis with baseline score as a covariate, fixed effects for sites, visits(categorical), treatment group and treatment group by visit interaction. The within patient covariance structure of the errors was assumed to have the most general structure, “unstructured” and the maximum likelihood method of estimation was used. Also based on this model, the difference was smaller at week 42:  $1.24 \pm .84$  S.E.,  $p=.144$  (see Table 13). There was a nominally significant interaction between site and treatment group in this model,  $p<0.0001$ . The sponsor came up with a post-hoc re-pooling of sites (8 pooled sites) which showed no site by treatment group interaction,  $p=0.1696$ . Alternatively, this reviewer found that replacing site with country (U.S. or Canada) in the model did not produce a significant treatment by country interaction and it also did not produce a nominally significant result for 2mg early vs. placebo/2mg delayed at week 52:  $-1.9439 \pm 1.19$ ,  $p=0.1039$ .

A two group analysis leaving out the 1mg group also had a significant site by treatment group interaction  $p=.0137$ . Averaged over the interaction the week 52 2mg early vs. placebo/2mg delayed difference was not significant  $-1.97 \pm 1.00$  S.E.,  $p=.0509$ . If we were to ignore the interaction and re-run the model without it the week 52 2mg early vs. placebo/2mg delayed difference was  $-2.0127 \pm 1.03$  S.E.,  $p=0.0515$ .

LOCF analysis, in which missing data at the time of interest is imputed with the last available post-baseline assessment, has been heavily criticized in the literature in the last few years and in such a case as this, a two phase study in a degenerative disease in which the early group may have a symptomatic advantage at the beginning of the 2<sup>nd</sup> phase it’s use would seem quite inappropriate. In particular, in a case with many early dropouts before the symptomatic effect is established in the delayed start group an LOCF imputation analysis could lead one if not careful into thinking that a purely symptomatic treatment was disease modifying.

The LOCF analysis also had the complication of a significant site by treatment group interaction,  $p=.0055$ . Averaged over site specific treatment difference estimates the 2mg early vs. placebo/2mg delayed difference was  $1.66$ ,  $p=.0587$ . The sponsor’s post hoc re-pooling of sites gave  $-1.83$ ,  $p=.0656$  when the 1mg group was excluded. Based on the original site pooling plan the model also didn’t have a significant interaction when the 1 mg group was excluded for the analysis and the estimated 2mg early vs. placebo/2mg delayed difference was  $-1.87$ ,  $p=.0347$ . However, there is no indication in the analysis plan or study report that the 2mg early vs. 2mg delayed comparison was planned to be performed excluding the 1 mg group.

The LOCF analysis adjusted for country instead of site yielded an estimated 2mg early vs. placebo/2mg delayed difference of  $1.88$ ,  $p=.0655$ .

Table 12 TEMPO Active Phase Observed Mean Changes in UPDRS by Last Assessment Time

Last Assessment Time	PLACEBO/2MG DELAYED			1MG/1MG			2MG/2MG		
	N	BASELINE UPDRS	CHANGE	N	BASELINE UPDRS	CHANGE	N	BASELINE UPDRS	CHANGE
		MEAN (SD)	Mean (SD)		Mean (SD)	Mean (SD)		Mean (SD)	Mean (SD)
32	30	32.3 (11.6)	8.2 (6.6)	19	28.1 (8.6)	5.6 (6.7)	29	31.5 (9.0)	2.0 (6.4)
42	10	25.5 (14.3)	5.1 (3.7)	15	28.3 (12.3)	3.2 (7.3)	13	26.7 (8.3)	4.8 (12.2)
52	91	20.9 (8.6)	2.9 (9.4)	89	23.0 (11.7)	2.3 (8.7)	81	23.2 (8.9)	1.6 (6.9)

Table 13 Repeated Measures Analysis of Group Difference By Week in Active Phase

GROUP VS PLACEBO/ WEEK	ESTIMATED GROUP DIFF	S. E.	P-VALUE
1MG Early Week52	-0.6729	1.0202	0.5101
1MG Early Week42	-1.5947	0.8253	0.0542
1MG Early Week32	-2.3349	0.7995	0.0037
2MG Early Week52	-2.0466	1.0399	0.0501
2MG Early Week42	-1.2366	0.8444	0.1441
2MG Early Week32	-2.3737	0.8100	0.0036

\*Although there was a significant site by group interaction it was assumed =0 in this model

With only three visits in the Active Phase we can't be too certain about the functional form between UPDRS Change and Time. However, one test for nonlinearity, comparing a saturated mean model to a linear model over all three groups for the active phase did not show enough evidence of nonlinearity to reject linearity,  $p=0.1386$ .

Based on a model of the Active Phase assuming a random intercept and slope for each subject with baseline score as a covariate and site adjustment (as specified in the SAP) the 2mg early slope was numerically bigger than the placebo/2mg delayed slope  $0.02060 \pm 0.045$  S.E.,  $p=0.6453$ . The 1 MG slope was also numerically bigger than placebo/2mg delayed  $0.08074 \pm 0.044$  S.E.,  $p=0.0657$ . For the next study, the sponsor prespecified a non-inferiority margin for the slope difference of 0.15. In this case, the confidence interval for the 1MG vs. placebo/2mg delayed slope difference would exceed that margin slightly and the 2mg early vs. placebo /2mg delayed slope difference would be slightly less than the margin, i.e., non-inferior. Note that this non-inferiority of slopes hypothesis was not conceived of before the TEMPO study so the foregoing slope analysis is all exploratory. It seems likely to this reviewer that a revisiting of this study's data like this was the basis for the slope margin proposed for the next study.

The estimated difference at week 52 between 2mg early and placebo /2mg delayed based on this random subject intercept and slope model was  $-1.7172 \pm 0.9954$  S.E.,  $p=0.0857$ . The estimated difference at week 52 between 1mg early and placebo /2mg delayed based on this random subject intercept and slope model was  $-0.7501 \pm 0.9739$  S.E.,  $p=0.4419$ .

When the analysis excluded the 1MG group the 2MG early slope was still numerically bigger than the placebo/2mg delayed slope  $0.02083 \pm 0.04542$   $p=0.6472$ . The estimated difference at week 52 based on this random subject intercept and slope model was  $-1.6628 \pm 1.0040$  S.E.,  $p=0.0995$ .

### 3.2.2 ADAGIO (Study 500)

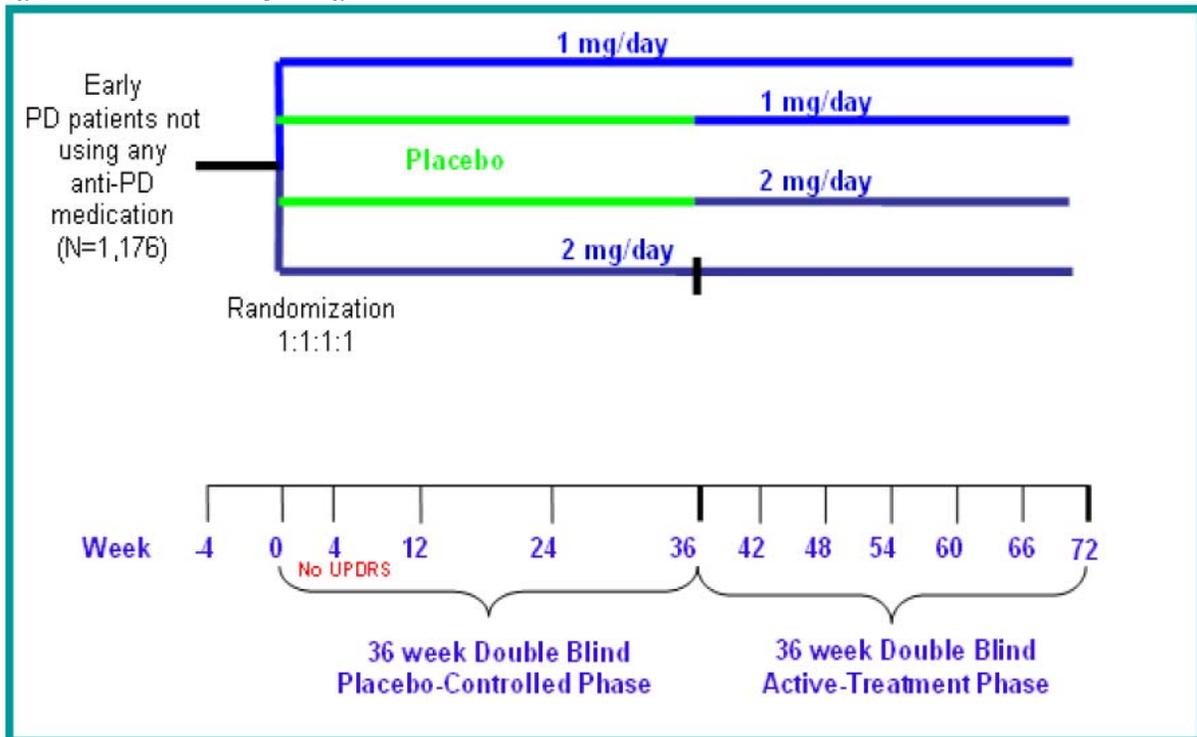
The first subject was enrolled on 4th November 2005 and the Last Subject Completed the Trial on 30th April 2008.

The primary objective of this study was to assess rasagiline as a disease modifying therapy in Parkinson’s disease.

#### 3.2.2.1 Study Design and Analysis Plan

This study was comprised of 2 phases: Phase I – a 36-week double-blind, placebo controlled Phase, and Phase II – a 36-week double-blind, active-treatment phase (see Figure 2).

Figure 2 ADAGIO Study Design



After being found eligible to participate in the study, subjects were to be allocated in a 1:1:1:1 ratio into one of the following four treatment groups based on a randomization scheme with blocks stratified by centers:

1. 1 mg/day rasagiline during phase I and phase II (1 mg early start)
2. 2 mg/day rasagiline during phase I and phase II (2 mg early start)
3. placebo during phase I followed by 1 mg/day rasagiline during phase II (1 mg delayed start)
4. placebo during phase I followed by 2 mg/day rasagiline during phase II (2 mg delayed start)

#### Phase I: Double-Blind Placebo-Controlled

Scheduled in-clinic visits were conducted at baseline and at weeks 4, 12, 24 and 36.

Unscheduled visits were conducted at any time to assess a subject's need for additional anti-PD therapy, for safety reasons or for any other reason. If, at any time (see Amendment No. 1) during the placebo-controlled phase, the investigator determined that a subject needed additional anti-PD therapy, the subject proceeded to Phase II of the study.

#### Phase II: Double-Blind Active-Treatment

Following the completion of Phase I or following a need for additional anti-PD therapy during Phase I, subjects transferred to Phase II. Based on the above randomization scheme, all subjects received active treatment (either 1 mg or 2 mg rasagiline) during this phase according to their original randomization allocation. If, at any time during the active-treatment phase, the investigator determined that a subject needed additional anti-PD therapy, the subject was prematurely withdrawn from the study.

Scheduled in-clinic visits were conducted at weeks 42, 48, 54, 60, 66 and 72. Unscheduled visits were conducted at any time to assess a subject's need for additional anti-PD therapy, for safety reasons or for any other reason.

### **Efficacy**

#### **UPDRS (Version 3)**

Subjects were to be assessed on Parts I, II and III of version 3 of the UPDRS at all scheduled study visits besides the screening and Week 4 visits. Part I

assesses the mental state of the subject in the week prior to the visit, Part II assesses the activities of daily living of a subject in the week prior to the designated visit and Part III assesses motor disabilities of a subject at the time of the visit. A total of 31 items are included in Parts I, II and III. Each item receives a score ranging from 0 to 4 where 0 represents the absence of impairment and 4 represents the highest degree of impairment.

The sum of Parts I, II and III at each study visit provides a Total UPDRS score.

Both the primary and secondary efficacy endpoints were to be based on changes from baseline in Total UPDRS scores.

**According to the protocol Subjects should be assessed on the UPDRS by the same investigator at all study**

**visits.** All study investigators were to receive instruction on how to complete the UPDRS and were, for study purposes, to be certified to perform the assessment. This was to ensure the standardization of the procedure.

### **SAMPLE SIZE RATIONALE**

The power was estimated under the assumption that the pooled standard deviation

of the mean change from baseline in total UPDRS during the active treatment phase is 6.3 UPDRS points and the pooled standard deviation of the slope is 0.35 UPDRS points per week, both estimated from trial TVP-1012/232 (TEMPO).

Results of these calculations show that a total of 935 subjects entering the active treatment phase (about 1100 subjects randomized, assuming 15% dropouts), will provide:

87% power for detection (at 5% significance level) of a statistical significant difference of a mean of 1.8 UPDRS points or more in the change from baseline in total UPDRS during the active-treatment phase between the two treatment groups.

99% power to detect (at 2.5% significance level due to adjustment for multiple comparisons, and non-inferiority threshold of 0.15 UPDRS points per week) non-inferiority between the slopes of the rasagiline 1 mg (2 mg) early-start group as compared to rasagiline 1 mg (2 mg) delayed-start group.

The original protocol was dated 4 April, 2005. Amendment 1, dated 31 May 2006, changed the assumed dropout rate from 25 to 15% and thereby increased the sample size from 800 to 935 for the active phase based on 1100 randomized (power increases from 80 to 87 %).

In amendment 2, dated 04 December 2007, the interim analysis planned for the time of 80% patient completion was cancelled as described below. *An interim analysis was planned in the protocol at the time point at which approximately 80% of the primary efficacy endpoint data would have been available. Due to the fact that recruitment of 1176 subjects exceeded pre-study projections, 80% of the subjects will complete the study in January 2008. At this time point the time difference between the interim analysis and study termination will be 4 months as the last subject is expected to finish the study in May 2008. Therefore the interim analysis will not be performed as originally planned and all analyses will be performed after study completion. This change will not affect safety or efficacy analyses except for the fact that the Lan-DeMets correction of type I error will no longer be required.*

The ITT Data Analysis Set was to consist of all subjects randomized with at least one post baseline measurement. In accordance with the ITT principle, subjects were to be kept in their originally assigned treatment group.

Amendment 2 also changed the Principal Efficacy Cohort, defined as all subjects entering the active phase who have at least 2 available UPDRS measurements during the active phase, to the Active Efficacy Data Analysis set (ACTE), defined as subjects entering the active phase with at least 24 weeks of treatment during the PC phase and at least one available Total UPDRS measurement during the active phase from week 48 onward.

The ITT analysis set was to serve for analysis of hypothesis #1 of the primary efficacy analysis, and secondary and additional efficacy endpoints for the PC phase.

The ACTE data analysis set was to serve for analysis of hypotheses #2 and #3 of the primary efficacy analysis, and secondary and additional efficacy endpoints for the active phase.

The Statistical Analysis Plan, version 2 (final), was dated 27 Feb, 2008.

### **Principal Efficacy Analysis**

#### **Primary Efficacy Endpoint**

The primary efficacy endpoint for this trial was to be the change from baseline in total UPDRS score (sum of parts I, II and III).

Observations of all subjects entering the active-treatment phase who have at least 2 available UPDRS measurements were to be analyzed. The UPDRS measurements at the end of the placebo-controlled phase prior to the active treatment phase were not to be included in the principal statistical analysis.

#### **Multiple Comparisons Adjustment**

The principal efficacy analysis consists of three hierarchical statistical hypothesis tests that were to be applied on the primary efficacy endpoint.

The Hochberg's Step-Up Bonferroni method for multiple comparisons between treatment groups (2 comparisons: early-start group compared to delayed-start group for Rasagiline 1 mg and 2 mg), in combination with the hierarchical method for the three hypotheses tests, are used to maintain the experiment-wise type I error of 5% according to the following procedure:

If the first null hypothesis comparing the early-start group to the delayed-start group is rejected for both rasagiline doses at an alpha level of 5% then no adjustment to the alpha level is performed and both comparisons are declared as statistically significant. If the first null hypothesis is not rejected for one of the doses at an alpha level of 5%, then the other dose is tested using an alpha level of  $5\%/2 = 2.5\%$ .

Each statistically significant dose, as determined by the test of hypothesis #1, is further tested for hypothesis #2 using an alpha level of 5% and the Hochberg's Step-Up Bonferroni method as described for hypothesis #1 above.

Each statistically significant dose, as determined by the test of hypothesis #2, is further tested for hypothesis #3 using an alpha level of 5% and the Hochberg's Step-Up Bonferroni method as described for hypothesis #1 above.

#### **Primary Efficacy Endpoint**

According to the final amended protocol and the SAP (see Amendment No.3) 3 hypotheses constitute the primary efficacy endpoint. All are based on changes from baseline in Total UPDRS scores (sum of parts I, II and III):

1) Hypothesis #1: Slopes Superiority of Rasagiline over Placebo in the PC Phase

H0: Slope(Rasagiline) – Slope(Placebo) = 0

HA: Slope(Rasagiline) – Slope(Placebo)  $\neq$  0

Where slope is the model estimate of the change from baseline in total UPDRS per week.

In this analysis, all available post baseline observations in the PC phase of the trial were to be analyzed (ITT data analysis set, weeks 12, 24 and 36).

The placebo groups for rasagiline 1mg and 2 mg were to be combined into one placebo group.

The statistical model was to be a Repeated Measures Mixed Linear Model with random intercept and slope (SAS MIXED procedure with RANDOM subcommand). The model was to include the following fixed effects: treatment group, continuous week in trial by treatment interaction, center and baseline Total UPDRS score. The individual subject intercept and the week effects were also to be included in the model as random effects. The unstructured covariance matrix between the intercept and slopes estimates was to be used.

Two comparisons were to be derived from the model: slope difference of rasagiline 1mg group from the placebo group and slope difference of rasagiline 2m group from the placebo group.

## **2) Hypothesis #2: Superiority of Early over Delayed Start at Week 72**

**H<sub>0</sub>: LSM(Early Start Group at Week 72) - LSM(Delayed Start Group at Week 72) = 0**

**H<sub>A</sub>: LSM(Early Start Group at Week 72) - LSM(Delayed Start Group at Week 72) ≠ 0**

(Where LSM is Least Square Means)

In this analysis, observations of all subjects entering the active phase with at least 24 weeks of treatment during the PC phase and at least one available Total UPDRS measurement during the active-treatment phase from weeks 48, 54, 60, 66 or 72, were to be analyzed (ACTE data analysis set).

The statistical model was to be a Repeated Measures model (SAS MIXED procedure with REPEATED sub-command). The model was to include the following fixed effects: categorical week in trial by treatment interaction, center, and baseline Total UPDRS score. The unstructured covariance matrix for repeated observations within subjects was to be used. In case that the model did not converge, a simpler covariance structure with less parameters was to be used, e.g. Heterogeneous Autoregressive(1) (ARH(1)) or Autoregressive(1) (AR(1)).

The LSM at week 72 of the change from baseline in Total UPDRS was to be compared between the rasagiline 1mg early-start group and the rasagiline 1mg delayed-start group and between the rasagiline 2mg early-start group and the rasagiline 2mg delayed-start group.

## **3) Hypothesis #3: Slopes Non-Inferiority of Early Start over Delayed Start in the Active Phase**

**H<sub>0</sub>: Slope(Early Start Group) - Slope(Delayed Start Group) > 0.15**

**H<sub>A</sub>: Slope(Early Start Group) - Slope(Delayed Start Group) ≤ 0.15**

In this analysis, observations of all subjects entering the active phase with at least 24 weeks of treatment during the PC phase and at least one available Total UPDRS measurement during the active-treatment phase from weeks 48, 54, 60, 66 or 72, were to be analyzed (ACTE data analysis set).

The statistical model was to be a Repeated Measures Mixed Linear Model with random intercept and slope (SAS MIXED procedure with RANDOM subcommand). The model was to include the following fixed effects: treatment group, continuous week in trial by treatment interaction, center and baseline Total UPDRS

score. The individual subject intercept and the week effects were also to be included in the model as random effects. The unstructured covariance matrix between the intercept and slopes estimates was to be used.

Non-inferiority test for the difference in slopes between the treatment groups was to be performed.

The one sided 95% Confidence Interval (CI) was to be calculated for the difference between the slopes of the rasagiline 1mg early-start group and the rasagiline 1mg delayed-start group and between the slopes of the rasagiline 2mg early-start group and the rasagiline 2mg delayed-start rasagiline group. The inferiority null hypothesis of the early-start group slope over the delayed-start group slope was to be rejected, if the upper limit of the one sided 95% CI for the difference in slopes did not cross the non-inferiority margin of 0.15 UPDRS points per week.

### **Test of Linearity for Hypothesis #3**

To test if the relation between the primary endpoint and week in trial is linear within treatment groups in the active phase, a Mixed Models Repeated Measures model (SAS MIXED procedure with REPEATED sub-command) was to be fitted, once with week effect as continuous variable (representing a linear relation), and second, with week effect as a categorical variable (representing a general relation). The model was to include the following fixed effects: week in trial by treatment interaction (once as continuous and once as categorical), center, and baseline Total UPDRS score. The unstructured covariance matrix for repeated observations within subjects was to be used. In case that the model did not converge, a simpler covariance structure with less parameters was to be used, e.g. Heterogeneous Autoregressive(1) (ARH(1)), Autoregressive(1) (AR(1)) or Compound Symmetry (CS). The -2 log likelihood ratio test for nested models, was to be used to test the contribution of the "general relation" model beyond the "linear relation" model. The hypothesis of linearity within treatment groups was to be rejected, if the p-value of the above test was less than 0.05.

### **Alternative Analysis in Case that the Hypothesis of Linearity within Treatment Groups for Hypothesis #3 is Rejected**

If the hypothesis of linearity within treatment groups for hypothesis #3 was rejected, the model stated for hypothesis #3 was to be replaced by a Mixed Model Repeated Measures model (SAS MIXED procedure with REPEATED sub-command) with week in trial as categorical (class) variable. The model was to include the following fixed effects: categorical week in trial by treatment interaction, center, and baseline Total UPDRS score. The unstructured covariance matrix for repeated observations within subjects was to be used. In case that the model did not converge, a simpler covariance structure with less parameters was to be used, e.g. Heterogeneous Autoregressive(1) (ARH(1)) or Autoregressive(1) (AR(1)).

The non-inferiority tests for the slopes difference between the early and delayed start treatment groups, were to be replaced by testing whether the average of the differences between the early and delayed start treatment groups at weeks 66 and 72 were not closer to zero than the average of the differences between the early and delayed start treatment groups at weeks 48 and 54. This was to be performed using a non-inferiority test on the estimate of the difference between the average of the

differences between the early and delayed start treatment groups at weeks 66 and 72 and the average of the differences between the early and delayed start treatment groups at weeks 48 and 54. The following formulae describes this estimate:

Estimate for non-inferiority test =

$$\frac{((LSM_{Early,W72} - LSM_{Delayed,W72}) + (LSM_{Early,W66} - LSM_{Delayed,W66}))/2 - ((LSM_{Early,W48} - LSM_{Delayed,W48}) + (LSM_{Early,W54} - LSM_{Delayed,W54}))/2}{2}$$

where  $LSM_{Early,W48}$  is the Least Square Mean of change from baseline in total UPDRS of the early start group at week 48, etc. The non-inferiority margin was to be 2.7 UPDRS units.

A non-inferiority margin of 2.7 UPDRS units was derived from the non-inferiority margin of 0.15 UPDRS units per week where linearity is assumed according to the following:

The time interval that corresponds to the difference between the average of observations at weeks 66 and 72  $((66+72)/2=69)$  and the average of observations at weeks 48 and 54  $((48+54)/2=51)$  is 18 weeks (week 69 - week 51). Hence the matching non-inferiority margin for the non-linear alternative is  $0.15 * 18 \text{ weeks} = 2.7 \text{ UPDRS units}$ .

In the original protocol there were different methods for testing for linearity and performing an alternative analysis in case of nonlinearity. These were deleted in favor of the new methods in amendment 3, dated 04 March 2008.

#### **Original Test of Linearity (Original protocol, dated 04 April 2005)**

In order to test if the relation between the primary endpoint, and week in trial is linear, an ANCOVA model with linear and quadratic week effects was to be fitted.

The contribution of the treatment by week squared effect over the linear model with the treatment by week effect was to be tested.

#### **Alternative Analysis in Case that Non-Linearity is Detected(original protocol, dated 04 April 2005)**

If the linearity in time (week) was rejected in favor of the quadratic model, then the test for non-inferiority of slopes was to be based on the computed slope at the last nominal evaluation time that is at 72 weeks. The slope was to be computed as the first derivative of the estimated quadratic equation:  $\beta_0 + \beta_1 * \text{Week} + \beta_2 * \text{Week}^2$ , where  $\beta_0$  denotes the intercept estimate,  $\beta_1$  the estimate of the coefficient of week, and  $\beta_2$ , the estimate of the coefficient of week squared for a given treatment group.

Hence, the slope at week 72 equals to:  $\beta_1 + 2 \beta_2 * (72) = \beta_1 + 144\beta_2$  for each group.

A test to compare these slopes between the early and delayed groups at each dose was then to be performed.

#### **Blinded Variance Estimate**

To examine whether the variance estimates that were used in the sample size calculations were adequate, a blinded assessment of the variance magnitude was to be performed after 1/3 of the subjects had completed the trial. A blinded evaluation of the early dropout rate was also to be performed at this time point. If the variance estimates were larger by 10% or more than those projected, or the early dropout rate



missing data imputation method: The mean of the change from baseline of the rasagiline 1mg/2mg delayed-start group, at the relevant weeks (48, 54, 60, 66 or 72), was to be used to impute the missing data of all subjects (regardless if assigned to delayed or early start groups), receiving delayed or early rasagiline 1mg/2mg.

3. A sensitivity analysis was to be performed by multiple imputation method (Rubin<sup>2</sup>, Little and Rubin<sup>3</sup>) using the SAS MI and MIANALYZE procedures. The multiple imputation was to be performed in 3 stages: First, the Markov-Chain-Monte-Carlo (MCMC) method was to be used on the UPDRS repeated measurement at weeks 0, 12, 24, 36, 42, 48, 54, 60, 66 and 72, to impute the non-monotone missing values and to achieve a monotone missing values structure. The number of imputations was to be 5. This procedure was to be done for each treatment group separately.

Second, on the output data from the first stage, the monotone imputation method (with one imputation) was to be used on the repeated UPDRS measurement at weeks 0, 12, 24, 36, 42, 48, 54, 60, 66 and 72, subject age, time to additional anti PD treatment need as continuous variables, and on treatment group, center and sex as categorical (class) variables. The regression method was to be used for all continuous variables and the discriminant function method was to be used for all class variables. Finally the model used for the principal analysis was to be activated on the imputed data set, separately for each one of the 5 imputations sets, yielding 5 parameter estimates sets, that were then to be combined and analyzed in the MIANALYZE procedure for appropriate parameter estimates and p-value calculations. The key sensitivity analysis (meaning that differences of conclusions from the principal analysis would lead to further investigation of the missing observations pattern) was to be the multiple imputation sensitivity analysis.

### **3.2.2.2 Subject Disposition**

Two of the subjects who underwent randomization to the 1 mg delayed-start group withdrew their decision to participate in the study prior to receiving even one dose of the study drug or having any post-baseline measurements. Consequently these 2 subjects are not counted in any study analyses, and the overall study ITT data analysis set comprises 1174 subjects.

For the 1174 subjects included in the study ITT data analysis set 298 subjects were randomized to the 1 mg delayed-start treatment group, 288 subjects to the 1 mg early-start treatment group, 295 subjects to the 2 mg delayed-start treatment group, and 293 subjects to the 2 mg early-start treatment group.

There was an overall premature termination rate of 18.7% that was similar between the 4 treatment groups, although slightly higher in each of the delayed groups compared to the early groups ( Table 14).

The most common reason for prematurely terminating the study, with an overall rate of 9.8%, was the need for additional anti-PD therapy.

**Table 14 Subject Disposition**

TVP1012/500 (ADAGIO)	1 mg Delayed Start		1 mg Early Start		2 mg Delayed Start		2 mg Early Start		All	
	N	%	N	%	N	%	N	%	N	%
<b>Received Double-Blind Medication</b>	<b>298</b>	<b>100.0</b>	<b>288</b>	<b>100.0</b>	<b>295</b>	<b>100.0</b>	<b>293</b>	<b>100.0</b>	<b>1174</b>	<b>100.0</b>
<b>Completed the Study</b>	<b>231</b>	<b>77.5</b>	<b>238</b>	<b>82.6</b>	<b>241</b>	<b>81.7</b>	<b>244</b>	<b>83.3</b>	<b>954</b>	<b>81.3</b>
<b>Prematurely Terminated the Study</b>	<b>67</b>	<b>22.5</b>	<b>50</b>	<b>17.4</b>	<b>54</b>	<b>18.3</b>	<b>49</b>	<b>16.7</b>	<b>220</b>	<b>18.7</b>

Copied from page 100 of sponsor's study report

The rate of early transfer from the placebo-controlled phase to the active-treatment phase was about 20% in each of the placebo groups (1 mg and 2 mg delayed-start treatment groups) compared to about 10% in each of the early-start treatment groups (Table 15).

As seen in Table 16, altogether 4.9% of subjects transferred early from the placebo-controlled phase to the active-treatment phase prior to week 24 (making them ineligible for Active Phase analysis according to the SAP) and 10.2% transferred early between Weeks 24 to 34. A total of 7.1% subjects prematurely terminated during the placebo-controlled phase; there were no statistically significant differences between the treatment groups.

**Table 15 Summary of Discontinuations Prior to Active Phase and Early Transfers to Active Phase**

	1 mg Delayed Start		1 mg Early Start		2 mg Delayed Start		2 mg Early Start		All	
	N	%	N	%	N	%	N	%	N	%
<b>Received Double-Blind Medication</b>	<b>298</b>	<b>100.0</b>	<b>288</b>	<b>100.0</b>	<b>295</b>	<b>100.0</b>	<b>293</b>	<b>100.0</b>	<b>1174</b>	<b>100.0</b>
<b>Entered into Active Phase After Completing Placebo Phase</b>	<b>211</b>	<b>70.8</b>	<b>245</b>	<b>85.1</b>	<b>216</b>	<b>73.2</b>	<b>242</b>	<b>82.6</b>	<b>914</b>	<b>77.9</b>
<b>Early Transfer from Placebo Phase to Active Phase</b>	<b>59</b>	<b>19.8</b>	<b>28</b>	<b>9.7</b>	<b>59</b>	<b>20.0</b>	<b>31</b>	<b>10.6</b>	<b>177</b>	<b>15.1</b>
<b>Premature Termination During Placebo-Controlled Phase</b>	<b>28</b>	<b>9.4</b>	<b>15</b>	<b>5.2</b>	<b>20</b>	<b>6.8</b>	<b>20</b>	<b>6.8</b>	<b>83</b>	<b>7.1</b>

Note: This table was copied from page 103 of sponsor's study report

**Table 16 Eligibility for Primary Analysis of Active Phase**

	1 mg Delayed Start		1 mg Early Start		2 mg Delayed Start		2 mg Early Start		All	
	N	%	N	%	N	%	N	%	N	%
Received Double-Blind Medication	298	100.0	288	100.0	295	100.0	293	100.0	1174	100.0
Enter Active Treatment Phase Following complete placebo phase	211	70.8	245	85.1	216	73.2	242	82.6	914	77.9
Early Transfer into Active Treatment Phase Prior to Week 24	20	6.7	12	4.2	15	5.1	10	3.4	57	4.9
Early Transfer into Active Treatment Phase between Weeks 24 - 34	39	13.1	16	5.6	44	14.9	21	7.2	120	10.2
Not Entering Active Treatment Phase	28	9.4	15	5.2	20	6.8	20	6.8	83	7.1

Note: This table was copied from page 103 of sponsor’s study report

As seen in Table 17 a total of 12.6% subjects prematurely terminated the study during the active controlled phase; there were no statistically significant differences between the treatment groups. The need for additional anti-PD therapy constituted the most common reason for prematurely terminating the active-treatment phase, with no noteworthy differences between the treatment groups.

**Table 17 Subject Disposition During Active Phase**

TVP1012/500 (ADAGIO)	1 mg Delayed Start (N=298)		1 mg Early Start (N=288)		2 mg Delayed Start (N=295)		2 mg Early Start (N=293)		All	
	N	%	N	%	N	%	N	%	N	%
Entered into Active Phase	270	100.0	273	100.0	275	100.0	273	100.0	1091	100.0
Completed the Study	231	85.6	238	87.2	241	87.6	244	89.4	954	87.4
Premature Termination during Active Phase	39	14.4	35	12.8	34	12.4	29	10.6	137	12.6

Note: This table was copied from page 106 of sponsor’s study report

### Subjects during Active-Controlled Phase

A total of 12.6% subjects prematurely terminated the study during the active controlled phase (see Table 18); there were no statistically significant differences between the treatment groups. The need for additional anti-PD therapy constituted the most common reason for prematurely terminating the active-treatment phase, with no noteworthy differences between the treatment groups.

Table 18 Reasons for Subject Premature Termination During Active Control Phase

TVP1012/500 (ADAGIO)	1 mg Delayed Start (N=298)		1 mg Early Start (N=288)		2 mg Delayed Start (N=295)		2 mg Early Start (N=293)		All	
	N	%	N	%	N	%	N	%	N	%
Entered into Active Phase	270	100.0	273	100.0	275	100.0	273	100.0	1091	100.0
Completed the Study	231	85.6	238	87.2	241	87.6	244	89.4	954	87.4
Subject Withdrew Consent	8	3.0	3	1.1	1	0.4	2	0.7	14	1.3
Protocol Violation	1	0.4	.	.	1	0.4	.	.	2	0.2
Need for Additional Anti-PD Treatment	26	9.6	26	9.5	25	9.1	22	8.1	99	9.1
Failed to Return / Lost to Follow-Up	.	.	.	.	1	0.4	1	0.4	2	0.2
Adverse Events	3	1.1	5	1.8	6	2.2	3	1.1	17	1.6
Death	.	.	1	0.4	.	.	.	.	1	0.1
Other	1	0.4	.	.	.	.	1	0.4	2	0.2

Note: This table was copied from page 106 of sponsor's study report

Study protocol dictated that subjects should have had 6 post-baseline UPDRS assessments during the active-treatment phase at weeks 42, 48, 54, 60, 66 and 72. Those subjects who completed at least 24 weeks of the placebo-controlled phase and who had at least one UPDRS measurement at or from the week 48 visit onwards comprise the ACTIVE Efficacy (ACTE) data analysis set upon which hypotheses #2 and #3 for the primary statistical analysis are based. A total of 85% of the

overall study ITT data analysis set (996 out of 1174 patients) met the criteria stated above and were therefore included in the ACTE cohort. There were minor differences between the treatment groups with a range of 80% in the 1 mg delayed-start group and 88% in the 2 mg early-start group. Note that the difference between the 1MG early and delayed groups proportions (Table 20) in the ACTE cohort is nominally significant,  $p=0.0196$ .

The most frequent reason for being excluded from the ACTE dataset was termination prior to week 48 but early conversion to active treatment was also important (Table 19).

**Table 19 Reasons for Exclusion from ACTE Dataset**

		1 mg Delayed Start		1 mg Early Start		2 mg Delayed Start		2 mg Early Start		All	
		N	%	N	%	N	%	N	%	N	%
Received Double-Blind Medication		298	100.00	288	100.00	295	100.00	293	100.00	1174	100.00
Not in ACTE	No post-baseline UPDRS measurement	3	1.01	2	0.69	2	0.68	3	1.02	10	0.85
	Terminated prior to Week 48	37	12.42	23	7.99	29	9.83	22	7.51	111	9.45
	Early conversion into Active Treatment Phase Prior to Week 24	20	6.71	12	4.17	15	5.08	10	3.41	57	4.86

Note: This table was copied from page 112 of sponsor's study report

**Table 20 Number of ITT Subjects Eligible for Active Phase Analysis**

	Treatment Group				
	1 mg Delayed Start	1 mg Early Start	2 mg Delayed Start	2 mg Early Start	All
<b>Expected Number of Subjects</b>	<b>298</b>	<b>288</b>	<b>295</b>	<b>293</b>	<b>1174</b>
<b>Actual Number of Subjects</b>	<b>238</b>	<b>251</b>	<b>249</b>	<b>258</b>	<b>996</b>
<b>% Actual out of Expected</b>	<b>79.87</b>	<b>87.15</b>	<b>84.41</b>	<b>88.05</b>	<b>84.84</b>

Note: This table was copied from page 112 of sponsor's study report

### 3.2.2.3 Baseline Demographic Characteristics

The great majority of subjects were Caucasian (97.7%). Approximately 60% of subjects in all 4 treatment groups were male. The mean age of subjects in this study was approximately 62 years in all treatment groups with a majority in the 55 to 65 (36%) and 65 to 75 (34%) age categories.

Subjects were in the early stages of their disease with a mean Total UPDRS score of approximately 20 units (see Table 21), and a mean Hoehn and Yahr score of 1.5 units.

**Table 21 Descriptive Statistics of Total UPDRS at Baseline -ITT Set**

	<b>UPDRS Total (Baseline)</b>				
	<b>1 mg Delayed Start</b>	<b>1 mg Early Start</b>	<b>2 mg Delayed Start</b>	<b>2 mg Early Start</b>	<b>All</b>
<b>N</b>	<b>298</b>	<b>288</b>	<b>295</b>	<b>293</b>	<b>1174</b>
<b>Mean</b>	<b>20.25</b>	<b>20.57</b>	<b>19.93</b>	<b>20.83</b>	<b>20.39</b>
<b>SD</b>	<b>8.76</b>	<b>8.40</b>	<b>8.11</b>	<b>8.80</b>	<b>8.52</b>
<b>Min</b>	<b>3.0</b>	<b>6.5</b>	<b>5.0</b>	<b>3.5</b>	<b>3.0</b>
<b>Median</b>	<b>19.0</b>	<b>19.0</b>	<b>19.0</b>	<b>19.5</b>	<b>19.0</b>
<b>Max</b>	<b>49.5</b>	<b>53.0</b>	<b>47.0</b>	<b>52.5</b>	<b>53.0</b>

Note: table copied from sponsor's study report page 118

There were minor differences between the treatment groups of the ACTE data analysis set (see Table 21) and the study ITT data analysis set (see Table 22).

**Table 22 Descriptive Statistics of Total UPDRS at Baseline -ACTE Set**

<b>TVP1012/500 (ADAGIO)</b>	<b>UPDRS Total (Baseline)</b>				
	<b>1 mg Delayed Start</b>	<b>1 mg Early Start</b>	<b>2 mg Delayed Start</b>	<b>2 mg Early Start</b>	<b>All</b>
<b>N</b>	<b>238</b>	<b>251</b>	<b>249</b>	<b>258</b>	<b>996</b>
<b>Mean</b>	<b>19.10</b>	<b>20.53</b>	<b>19.24</b>	<b>20.27</b>	<b>19.80</b>
<b>SD</b>	<b>8.07</b>	<b>8.45</b>	<b>7.87</b>	<b>8.45</b>	<b>8.23</b>
<b>Min</b>	<b>3.0</b>	<b>6.5</b>	<b>5.0</b>	<b>3.5</b>	<b>3.0</b>
<b>Median</b>	<b>17.8</b>	<b>19.0</b>	<b>18.0</b>	<b>19.0</b>	<b>18.5</b>
<b>Max</b>	<b>49.0</b>	<b>53.0</b>	<b>47.0</b>	<b>51.5</b>	<b>53.0</b>

Note: Table copied from page 119 of sponsor's study report

**Reviewer’s Comment:** Actually, although the sponsor calls the differences minor, the p-value for the 1mg D vs. E comparison is .0563 based on ANOVA (.0523 for Wilcoxon rank sum test) so the difference in ACTE may be important (in the ITT set it was much bigger p=0.678). This is more of a concern for analysis of the active phase which may have more tendency to be a biased sample due to non-random dropouts because subjects had to make it through 48 weeks to be eligible for the active phase analysis. Also, if we combine the two early groups and the two delayed groups the baseline UPDRS Total means are different in ACTE: 20.39 (early) and 19.17 (delayed), p=0.0186. The 2mg difference: 20.27(early) vs. 19.24 (delayed) was numerically in the same direction as for 1mg but not nominally significant, p=0.157.

This reviewer also found several other nominally significant differences between IMG Early and Delayed Groups in the ACTE set that were not significant in the larger ITT set: Caucasian vs. Non-Caucasian; PIGD (which consists of the sum of 5 items: 3 UPDRS ADL subscale items (#13, 14, 15) and 2 UPDRS Motor subscale items (29 and 30) which seem to deal with gait and “freezing” while walking); UPDRS Motor subscale, UPDRS Brady Subscales [sum of the UPDRS Motor items (23, 24, 25, 26, 31)]. Such differences were not seen between the 2 mg early and 2mg delayed groups. The absolute difference in proportion non-Caucasian between treatment groups is small (3%) but because the overall proportion of non-Caucasian is so small the odds ratio of non-Caucasian between the two treatment groups is relatively big:  $4.2/95.8 * 98.8/1.2 = 3.6$ , i.e., the odds of being non-Caucasian are more than 3.5 times greater in the IMG Delayed group than the odds in the IMG Early group, p=0.049.

**Table 23 Potentially Important Race Difference between the Treatment Groups by Trial Phase**

NON-CAUCASIAN	1MG D	1MG E	FISHER’S EXACT P-VALUE
ACTE	10/238 (4.20%)	3/251 (1.20%)	.0490
ITT	10/295 (3.39%)	6/286 (2.10%)	.4490

**Table 24 Potentially Important Differences between the Treatment Groups in Baseline Variables by Trial Phase**

BASELINE VARIABLE	POP	1MG DELAYED						1MG EARLY						T-TEST PVALUE
		N	MEAN	STD	MEDI AN	MI N	MAX	N	MEAN	STD	MEDI AN	MI N	MAX	
UPDRS PIGD* subscale	ACTE	238	0.95	0.95	1.00	0.00	4.50	251	1.16	1.07	1.00	0.00	4.00	.024
	ITT	295	1.09	1.03	1.00	0.00	5.00	286	1.19	1.07	1.00	0.00	5.00	.256
UPDRS MOTOR subscale	ACTE	238	13.30	5.94	12.25	2.00	32.00	251	14.45	6.32	14.00	4.00	37.00	.039
	ITT	295	14.10	6.47	13.00	2.00	37.00	286	14.52	6.36	14.00	2.50	37.00	.428
UPDRS BRADY subscale	ACTE	238	5.91	3.19	5.00	0.00	16.00	251	6.55	3.39	6.00	0.00	17.00	.033
	ITT	295	6.27	3.51	5.50	0.00	21.50	286	6.58	3.41	6.00	0.00	17.00	.274
UPDRS Total	ACTE	238	19.10	8.07	17.75	3.00	49.00	251	20.53	8.45	19.00	6.50	53.00	.056
	ITT	295	20.29	8.77	19.00	3.00	49.50	286	20.59	8.42	19.00	6.50	53.00	.679

The median time from PD diagnosis until entry into the study ranged between 71.0 and 89.0 days across the treatment groups as seen in Table 25.

Table 25 Descriptive statistics of Time from PD Diagnosis (ITT)

<b>TVP1012/500 (ADAGIO)</b>		<b>1 mg Delayed Start</b>	<b>1 mg Early Start</b>	<b>2 mg Delayed Start</b>	<b>2 mg Early Start</b>	<b>All</b>
<b>Time from Diagnosis (days)</b>	<b>N</b>	<b>298</b>	<b>288</b>	<b>295</b>	<b>293</b>	<b>1174</b>
	<b>Mean</b>	<b>131.8</b>	<b>139.4</b>	<b>139.6</b>	<b>139.5</b>	<b>137.6</b>
	<b>SD</b>	<b>139.9</b>	<b>144.0</b>	<b>139.2</b>	<b>141.2</b>	<b>140.9</b>
	<b>Min</b>	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>
	<b>Median</b>	<b>71.0</b>	<b>79.5</b>	<b>85.0</b>	<b>89.0</b>	<b>83.0</b>
	<b>Max</b>	<b>547.0</b>	<b>540.0</b>	<b>530.0</b>	<b>546.0</b>	<b>547.0</b>

Note: This table was copied from the sponsor's study report, page 118

### 3.2.2.4 Sponsor's Results

#### 3.2.2.4.1 Placebo Controlled Phase

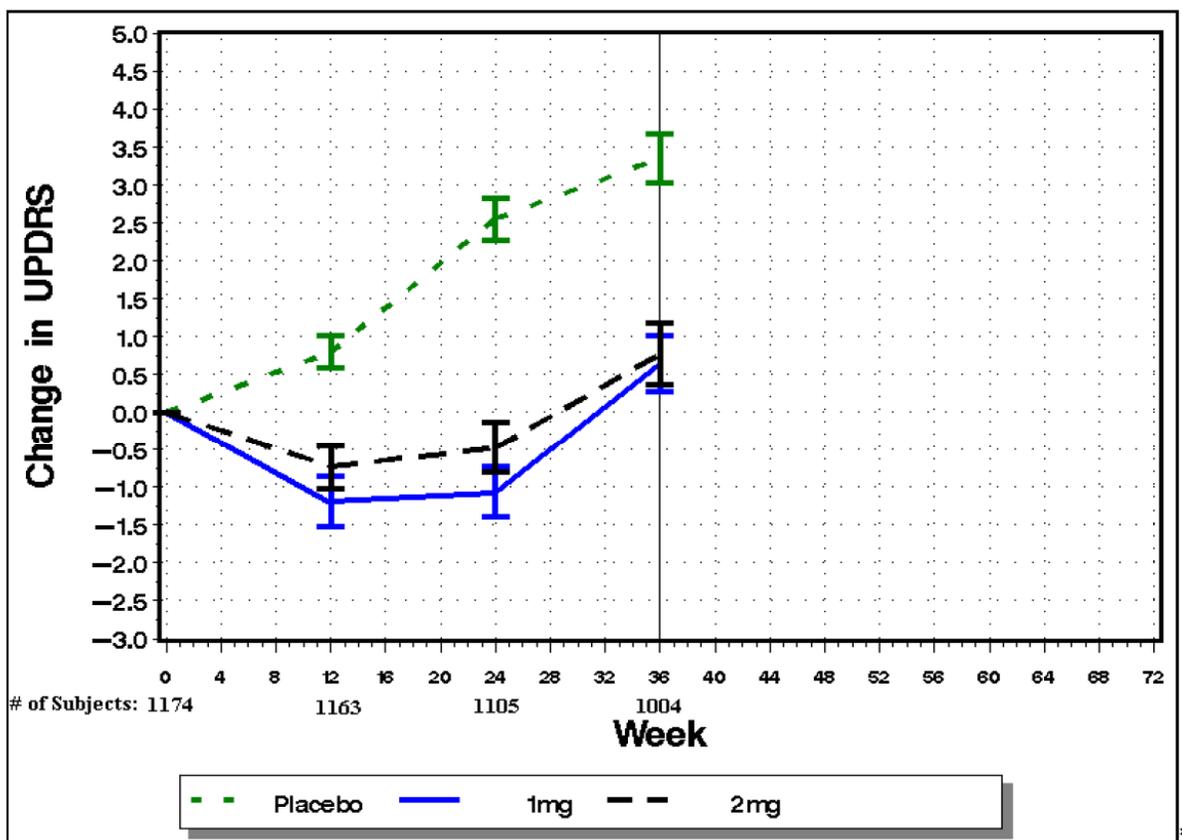
A larger rate of deterioration in UPDRS was evident in the placebo treatment group (0.139 UPDRS units/week) compared to both the 1 mg and 2 mg rasagiline early-start treatment groups (0.093 and 0.066 UPDRS units/week, respectively).

A point estimate of -0.046 UPDRS units/week was obtained for the difference between the 1 mg early-start rasagiline and placebo groups in the rate of UPDRS deterioration with 95% confidence intervals of -0.083 to -0.010. The p-value of 0.0133 was statistically significant at an alpha level of 5% (see Table 27).

A point estimate of -0.072 UPDRS units/week was obtained for the difference between placebo and the 2 mg early-start rasagiline group in the rate of UPDRS deterioration with 95% confidence intervals of -0.109 to -0.036. The p-value of 0.0001 was statistically significant at an alpha level of 5% (see Table 27).

Thus, summarizing the results from hypothesis #1 the null hypotheses were rejected for both the placebo to 1 mg rasagiline early-start comparison and the placebo to 2 mg early-start comparison.

Figure 4 Change from Baseline in Total UPDRS Score during Placebo-Controlled



Numbers represent the number of subjects per week

Note: This Figure was copied from page 122 of Sponsor's study report

Table 26 shows the estimated slopes of UPDRS change from baseline over time in the placebo controlled phase.

Table 26 Model Estimates for Changes per Week (Slope) in Total UPDRS during PC Phase Efficacy ITT set

		Estimate	SE	P-Value	Lower 95% CI Limit	Upper 95% CI Limit
<b>Data Analysis Set</b>	<b>Estimate</b>					
ITT	Placebo Slope	0.139	0.011	<.0001	0.117	0.160
	1 mg Slope	0.093	0.015	<.0001	0.063	0.122
	2 mg Slope	0.066	0.015	<.0001	0.037	0.096

Repeated Measures Mixed Linear Model with Random Intercept and Slope using Unstructured Covariance Matrix between Intercept and Slope Estimates

Note: This table was copied from page 125 of sponsor's study report

Table 27 shows the differences between early rasagiline and combined placebo in the estimated slopes of UPDRS change from baseline over time in the placebo controlled phase.

**Table 27 Comparison of Changes per Week (Slope) for Rasagiline vs. Placebo during PC Phase (week 12-Week 36) Efficacy ITT Set**

Data Analysis Set	Comparison	Estimate	SE	P-Value	Lower 95% CI Limit	Upper 95% CI Limit
ITT	1 mg-Placebo Slope Difference	-0.046	0.019	0.0133	-0.083	-0.010
	2 mg-Placebo Slope Difference	-0.072	0.019	0.0001	-0.109	-0.036

Repeated Measures Mixed Linear Model with Random Intercept and Slope using Unstructured Covariance Matrix between Intercept and Slope Estimates  
 Note: This table was copied from page 125 of sponsor's study report

### **Alternative Categorical Analysis, Efficacy ITT Data Analysis Set**

The changes from baseline in UPDRS during the placebo-controlled phase (based on only 3 possible post-baseline visits at Week 12, Week 24 and Week 36) demonstrated a statistically significant deviation from linearity ( $p=0.0001$ , Likelihood Ratio Test). This calls into question the results just described since they were based on a model assuming linearity. Thus to support the primary hypothesis #1 analysis the alternative categorical analysis was conducted by the sponsor post-hoc. The placebo group had a higher increase in UPDRS from Week 12 to Week 36 than either of the two rasagiline treated groups. A point estimate of -1.043 UPDRS units was obtained for the difference between the 1 mg early-start rasagiline and placebo groups in the difference between weeks 36 and 12 with a 95% confidence interval of -1.918 to -0.168. The p-value of 0.0196 was statistically significant at an alpha level of 5%.

A point estimate of -1.668 UPDRS units was obtained for the difference between the 2 mg early-start rasagiline and placebo groups in the difference between weeks 36 and 12 with a 95% confidence interval of -2.541 to -0.794. The p-value of 0.0002 was statistically significant at an alpha level of 5%.

### **Reviewer's Comments**

*Although the prespecified analysis of hypothesis 1 was significant for both early rasagiline groups the apparent increase in slope for these groups after week 24 (see Figure 4) is a cause for concern and may invalidate the primary analysis. A Slope comparison during placebo phase using actual time of assessment rather than classified week yielded the following.*

*ITT set: Plac vs. 1MG Early slope difference= 0.044,  $p=.0220$*

*Restricted to ACTE subgroup:*

*Combined Plac vs. 1MG Early slope diff= .035  $p=.0542$*

*1MG Del vs. 1MG Early slope diff= .036  $p=.0927$*

*As an alternative approach to deal with nonlinearity this reviewer also fit a repeated measures model (MMRM) followed by a contrast on the visit specific means to check for a quadratic fit. A quadratic fit to the estimated mean profile obtained from an MMRM model yielded a significant quadratic term for the 1 MG early group  $p=0.0002$ . Testing the 1MG early vs. combined placebo group difference in the resulting quadratic term, (which is a linear combination of the fitted visit means) gave  $p<.0001$ . Also, if we evaluate the slope of the fitted quadratic at week 36 we find a*

group difference in it favoring placebo  $\beta_P - \beta_{1E} = -0.1482$ ,  $p = .0037$ . Difference in slopes within the 1 MG early group over weeks 12 to 24 and 24 to 36, respectively, is estimated to be  $\beta_{24to36} - \beta_{12to24} = .1435$  (an increase in slope),  $p = .0002$ . For placebo the corresponding result is  $\beta_{24to36} - \beta_{12to24} = -0.0483$ ,  $p = .0845$ .

The estimated difference in group means between IMG Early and placebo at week 36 is 3.03 +/- .488 S.E. as compared to 3.66 +/- .429 S.E. at week 24. The estimated difference in group means between 2MG Early and placebo at week 36 is 3.19 +/- .476 S.E. as compared to 3.13 +/- .415 S.E. at week 24.

In the ACTE subgroup the estimated difference in group means between IMG Early and placebo at week 36 is 2.58 +/- .469 S.E. as compared to 3.24 +/- .410 S.E. at week 24.

In the ACTE subgroup the estimated difference in group means between 2MG E and placebo at week 36 is 2.52 +/- .465 S.E. as compared to 2.50 +/- .407 S.E. at week 24. So, in the ACTE subset while there is a nominally significant difference in means at week 36, the slope difference which was the parameter for hypothesis 1, is not nominally significant for IMG Early vs. combined placebo.

A repeated measures model with linear and quadratic effect for each treatment group (without subject random effects but allowing for correlation within patient through the SAS Repeated statement) should be as justifiable as the typical MMRM saturated model since they both have the same number of degrees of freedom (d.f.) for the model of UPDRS change as a function of time. An F test with 3 numerator d.f. for any nonzero quadratic terms among the treatment groups based on this model has a p-value of 0.0001 (also a likelihood ratio test between quadratic and linear models gives  $p = .0001$ ). This suggests again that the linear model is not adequate.

In conclusion, the slope is not well defined for week 12 and beyond in the placebo controlled phase because it appears to change with time as in a quadratic or other nonlinear model. While, there appears to be a difference in mean changes at week 36 this doesn't exactly address hypothesis 1 and the difference appears to be diminishing compared to week 24.

### **3.2.2.4.2 ACTIVE PHASE**

#### **3.2.2.4.2.1 Hypothesis #2: Mean Difference at Week 72/Active Week 36**

##### **Primary Analysis**

The model that was predefined in the SAP for hypothesis #2 was to have been applied on a combined 1 and 2 mg rasagiline data base, as it assumed no interaction of dose levels with the covariates in the model, baseline UPDRS and center. An underlying assumption for the simultaneous estimation of the treatment effect of the two doses (1 mg and 2 mg) is that the effect of the adjusting covariates in the model (baseline UPDRS and center) is the same for both dose levels. Following code breaking it was apparent that the pattern of response of the two dose levels in the Active phase was different and therefore the underlying assumption was tested using

the original model for hypothesis #2, with the addition of dose level (1 mg or 2 mg) by baseline UPDRS interaction and dose level by center interaction, applied on the combined dataset of four treatment arms. Indeed, both interactions were found to be important ( $p=0.0481$  and  $p=0.0125$  respectively). Since the analysis of the treatment effects is affected by these interactions, it was considered reasonable in the sponsor's opinion to adapt the final analysis. Estimation of the treatment effect of 1mg and 2 mg, in the presence of the above interactions, could be addressed by either addition of interaction terms to the model or by performing separate datasets analysis employing the original model structure. The separate data sets strategy was employed as the post-hoc primary analysis as it preserved the original model structure defined in the SAP and the results are presented below.

A higher deterioration from baseline to Week 72 was evident for the 1 mg delayed-start treatment group (4.495 UPDRS units) compared to the 1 mg early-start treatment group (2.815 UPDRS units). A point estimate of -1.680 UPDRS units was obtained for the difference between the 1 mg early-start and 1 mg delayed-start treatment groups in the change from baseline in UPDRS with 95% confidence interval of -3.15 to -0.21, and a p-value of 0.0250 (Table 28). The 2 mg delayed-start treatment group deteriorated from baseline to Week 72 by 3.111 UPDRS units, similarly to the deterioration of the 2 mg early-start treatment group (3.467 UPDRS units). A point estimate of 0.356 UPDRS units was obtained for the difference between the 2 mg early-start and 2 mg delayed-start treatment groups in the change from baseline in UPDRS with 95% confidence interval of -0.99 to 1.70. The p-value of 0.6028 was not statistically significant at an alpha value of 5%.

The comparison of the 1 mg delayed-start to the 1 mg early-start with a p-value of 0.0250 was statistically significant at an alpha level of 2.5% dictated by the method of adjustment for multiple comparisons if one accepts the revised primary analysis (post-hoc use of separated IMG data set).

A larger deterioration from baseline to Week 72 was evident for the 1 mg delayed start group (4.42 UPDRS units) compared to the 1 mg early-start group (2.996 UPDRS units) when the analysis was performed on a combined data set. A point estimate of -1.425 UPDRS units was obtained for the difference between the 1 mg early-start and 1 mg delayed-start treatment groups in the change from baseline in UPDRS with a 95% confidence interval of -2.85 to 0.004, and p-value of 0.0506 (see Table 28).

When the interaction terms of dose level by baseline UPDRS and dose level by center were added to the model, results derived from the combined data set were similar to those derived from the separate data sets (treatment effect: -1.606, p-value: 0.0191). As previously stated, it is because of these interactions that the analysis of hypothesis #2 was applied to separate data sets.

**Reviewer’s Comment:** *Note that a significant interaction suggests that at best the treatment effect is significantly inconsistent across the levels of the other non-treatment variable involved in the interaction, which implies that the mean of the treatment effect across that variable is not a broadly applicable quantity to the general study population. In the case of site/hospital, for example, this suggests that the treatment effect could depend on the hospital of administration and/or that there was significant variability in the practice of the different hospitals or their UPDRS assessment practices.*

Note that the separated data set for 1mg delayed and 1 mg early had 489 subjects and the separated data set for 2 mg delayed and 2mg early had 507 subjects. Thus, the combined set has 996 patients.

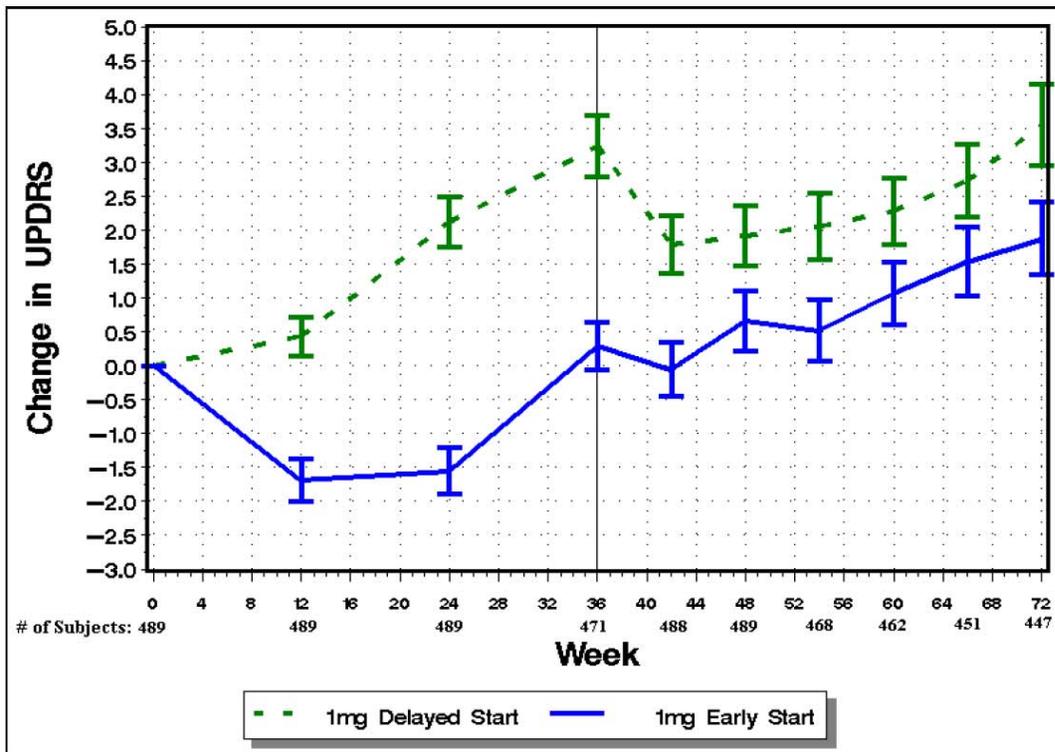
**Table 28** Dependence of Week 72 Difference on Analysis Data Set Used

ANALYSIS DATASET	GROUP	ESTIMATE DIFFERENCE	STD. ERROR	DF	T VALUE	PR >  T	LOWER	UPPER
Combined Dataset	1MG Early-1MG Delayed	-1.4245	0.7278	869	-1.96	0.0506	-2.8530	0.003982
	2MG Early-2MG Delayed	0.1791	0.7117	861	0.25	0.8014	-1.2178	1.5760
Separate 1MG	1MG Early-1MG Delayed	-1.6799	0.7469	430	-2.25	0.0250	-3.1478	-0.2119
Separate 2MG	2MG Early-2MG Delayed	0.3565	0.6845	406	0.52	0.6028	-0.9891	1.7020

Suppose that 1MG and 2MG early groups have the same effect of -1.6799 compared to delayed start. The probability that the maximum of two normally distributed variables identically distributed with a mean of -1.6799 and standard deviation of .7469 is greater than .3565 =  $1 - \Phi^2[(.3563 - -1.6799)/.7469] = .0064$ . Thus, it seems unlikely that the true 2mg delayed-early effect is the same as the observed 1mg result based on the separate data sets.

Figure 5 show the patterns of the mean change in UPDRS from baseline for the 1 MG groups over the course of the trial within the ACTE dataset.

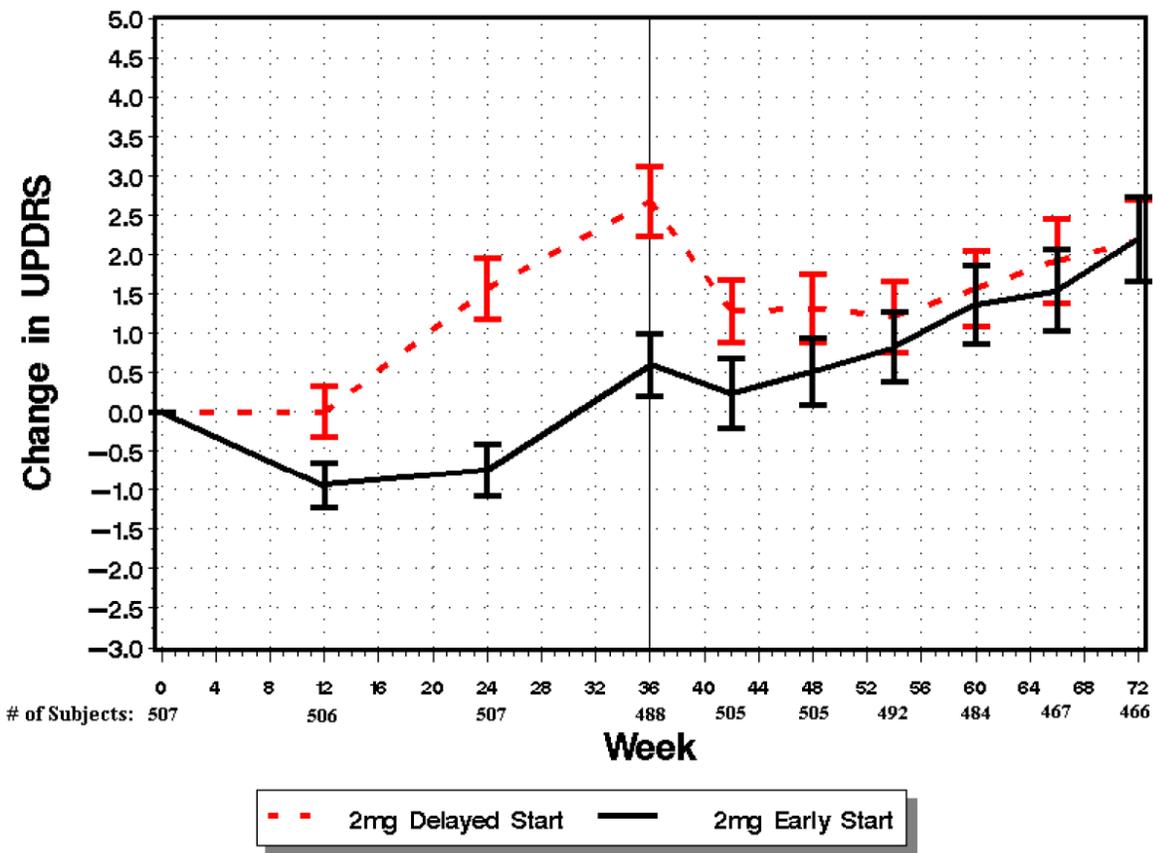
Figure 5 Mean UPDRS Change from Baseline Over Trial in 1 MG Groups (ACTE dataset)



\* Numbers represent the number of subjects per week. The larger number of subjects in Week 42 compared to Week 36 is due to the early transfer of subjects from Phase I to Phase II.

Figure 6 shows the patterns of the mean change in UPDRS from baseline for the 2 MG groups over the course of the trial within the ACTE dataset.

Figure 6 Mean UPDRS Change from Baseline Over Trial in 2 MG Groups (ACTE dataset)



\* Numbers represent the number of subjects per week. The larger number of subjects in Week 42 compared to Week 36 is due to the early transfer of subjects from Phase I to Phase II.

**Pooled Dose Analysis (1 mg combined with 2 mg), ACTE Data Analysis Sets**

As a sensitivity analysis we can pool the 1mg and 2mg doses. A larger deterioration from baseline to Week 72 was seen for the pooled delayed-start treatment group (3.623 UPDRS units) compared to the pooled early start treatment group (2.934 UPDRS units). The difference between these values, however, was not statistically significant (see Table 29).

Table 29 Comparison of Changes from Baseline to Week 72 in Total UPDRS Score, 1 mg Combined with 2 mg, ACTE Data Analysis Set

	Estimate	SE	P-Value	Lower 95% CI Limit	Upper 95% CI Limit
<b>Comparison</b>					
<b>Rasagiline Early-Delayed Start at week 72</b>	<b>-0.689</b>	<b>0.451</b>	<b>0.1268</b>	<b>-1.573</b>	<b>0.196</b>

Repeated Measures Analyses of Change from Baseline in Total UPDRS Score  
Hypothesis #2 for Combined Rasagiline 1 mg and 2 mg - Difference between Rasagiline 1 mg and 2 mg Early and Delayed Treatment Start at week 72 of the Active Phase  
Model with Autoregressive Covariance Matrix

Note: copied from sponsor’s study report page 140

The sponsor may have reported the pooled analysis results based on the different autoregressive covariance structure (noted below the table) because SAS Proc MIXED did not converge for the pooled analysis when the unstructured covariance was assumed within patient. However, this reviewer found that when starting values obtained from the primary analysis model fit of hypothesis 2 were used for the covariance parameters in SAS the model did converge. The effect size was slightly smaller and p-value was slightly bigger when the covariance structure was this most general “unstructured”, as in the primary analysis. In particular, the difference was -0.598, p=0.2410.

**Reviewer’s Assessment of Baseline UPDRS score by Treatment Group Interaction**

The treatment group by baseline UPDRS score interaction was significant in both the combined 1 and 2 mg dataset (p=0.0187) and in the separate 1mg dataset. This suggests that at least quantitatively the coefficient of the baseline UPDRS score covariate in the model of the change from baseline in UPDRS varied significantly by treatment group. If this is accounted for in the model then it means that there isn’t a broadly applicable treatment difference but rather the treatment difference depends on the baseline UPDRS score in the form of a linear equation.

Table 30 Baseline UPDRS score by Treatment Interaction: Baseline score coefficient estimates

TREATMENT GROUP	BASELINE UPDRS COEFFICIENT ESTIMATE	S. E.	DIFFERENCE FROM 0 P-VALUE
1 mg delayed	+0.1328	0.05021	0.0083
1 mg early	-0.04631	0.04677	0.3223
2 mg delayed	-0.05179	0.05143	0.3142
2 mg early	-0.1146	0.04625	0.0134

If we remember that the dependent variable is change from baseline which involves  $-1 \times \text{baseline score}$  then after accounting for the  $-1$  the various effects of baseline by group on the UPDRS Total score all have the same (+/-)sign (1+ the corresponding coefficient in the table) for the coefficient of the Baseline UPDRS score. One can also see this same result if one examines the results after changing the dependent variable to UPDRS Total score instead of UPDRS score-Baseline UPDRS score, but still including the Baseline UPDRS as a covariate in the model. Thus, because the signs are the same the baseline score by treatment group interaction is a quantitative interaction not a qualitative one. Since there is still a quantitative interaction between baseline score and treatment group in the 1MG only separate dataset this interaction problem attributed to the combined dataset doesn't seem to this reviewer like adequate justification for changing from the combined to the separate datasets.

Note that covariate effect estimates such as baseline UPDRS or Site effect estimates changing between the combined and the separate datasets might also be responsible for the slight changes in the respective treatment effect estimates. The best we can say is that based on the prespecified primary analysis the 1 mg result is not robustly positive at the multiplicity adjusted significance level of 0.025.

This reviewer also notes that apparent differences in the effect of the covariate Baseline UPDRS score between the groups may be influenced by the fact that the baseline score was nearly imbalanced between 1MG Early (Mean=20.53) & 1MG Delayed (Mean= 19.10) in the ACTE group. Separating out the 1 MG groups data from the 2 MG group's data, does not correct this potentially serious imbalance problem.

The average 1mg treatment difference (delayed-early) over the four baseline UPDRS quartiles subgroups is  $1.66 \pm .755$  S.E.,  $p=0.0283$ . A test of whether the quartile specific treatment effects are inconsistent has a p-value of 0.2810 (test for baseline quartile by treatment interaction), which suggests that they are reasonably consistent so that it makes sense to compute the above average. Because the p-value for the average is greater than 0.025, the multiplicity adjusted significance level, this too suggests that the "primary" analysis result based on the separate datasets is not robust.

### **Reviewer's Assessment of Pooled Site by Treatment Group Interaction**

As noted by the sponsor there was also a statistically significant Site by treatment group interaction, suggesting that site specific treatment group differences were significantly variable. There were 98 pooled site parameters included in the primary analysis model as pre-specified in the statistical analysis plan. Another post-hoc way to deal with the significant Site by treatment group interaction is to include Country in the model instead of Site. After adjusting for Country in the model instead of Pooled Site (Hosp1) we find using the separate 1MG data set that the difference between 1MG delayed and 1MG early groups at Week 72 becomes  $1.36 \pm .793$  S.E.,  $p=.0873$  (country\*TRT group interaction was not significant,  $p=.6861$ , so left out of model). The corresponding result for the 2mg separate analysis was  $-0.2438 \pm 0.7442$  S.E.,  $p=.7433$ . Using the combined data set we find the following.

	Estimate	S.E.	D.F.	t stat	p-value
1MG Delayed-Early Week 72	1.2197	0.7791	919	-1.57	0.1178
2MG Delayed-Early Week 72	-0.1653	0.7626	913	0.22	0.8284

The magnitude of the 1MG difference at Week 72 based on the Country adjusted analysis is noticeably lower than that based on the Site adjusted analysis (1.22 vs. 1.68).

In the U.S. subset (N=157 32% of 489) of the 1MG separate data set we find an estimate of +2.02, p=.1504 for the 1 MG delayed – 1mg Early 72 week difference. The corresponding 2mg difference in the U.S. subset was -1.32 +/- 1.21 S.E., p=0.2783. Note that the delayed 2mg group is numerically better, but not significantly better, here at week 72 as it was overall in the ACTE data set. These analyses were adjusted for sites within the U.S. subgroup. Note, however, that the site by treatment group interaction was significant (F test p=.0007), as it was overall, but ignored for this analysis in the U.S. subgroup of the 1MG separate dataset. Thus, there was significant variation of site specific treatment differences, i.e., inconsistency, between 1 mg early and delayed groups even within the U.S subset.

### **Assessment of Impact of Missing Data on Group Difference at Week 72 (Hypothesis 2)**

There was only about 9% missing Week 72 data within the ACTE dataset, however, relative to the ITT dataset there was 139/586= 24% for the 1MG groups. For the 2 mg groups there was 41/507 8% missing in the ACTE dataset and 122/588=21% in the ITT dataset.

### **CO and PP Data Analysis Sets for Separate Data Sets**

As for the ACTE data analysis set, the sponsor noted that larger deteriorations from baseline at Week 72 were evident in the 1 mg delayed-start groups compared to the 1 mg early-start groups of the CO and PP data analysis sets, with treatment effects in favor of rasagiline early treatment (Table 31). Similar deteriorations from baseline at Week 72 were evident for the 2 mg early and 2 mg delayed-start groups.

Table 31 Comparison of Changes from Baseline to Week 72 in Total UPDRS Scores – CO and PP Data Analysis Sets

		Estimate	SE	P-Value	Lower 95% CI Limit	Upper 95% CI Limit
Data Analysis Set	Treatment Group					
CO	1 mg Delayed Start	3.502	0.590	<.0001	2.342	4.662
	1 mg Early Start	1.910	0.522	0.0003	0.885	2.936
	2 mg Delayed Start	2.551	0.508	<.0001	1.552	3.549
	2 mg Early Start	2.463	0.494	<.0001	1.491	3.434
PP	1 mg Delayed Start	3.514	0.597	<.0001	2.340	4.688
	1 mg Early Start	1.920	0.524	0.0003	0.889	2.951
	2 mg Delayed Start	2.604	0.507	<.0001	1.608	3.600
	2 mg Early Start	2.581	0.495	<.0001	1.607	3.554

Mixed Model Repeated Measures Model with categorical week in trial using Unstructured Covariance Matrix  
 Separate analysis by each dose level

Note: This table was copied from sponsor’s study report page 133

The 1MG difference at week 72 in the completers set was not significant at the multiplicity adjusted significance level of 0.0250 (Table 32).

Table 32 Group Differences in Changes from Baseline to Week 72 in Total UPDRS Scores – CO and PP Data Analysis Sets

			Estimate	SE	P-Value	Lower 95% CI Limit	Upper 95% CI Limit
Data Analysis Set	No. of Subjects	Comparison					
CO	409	1 mg Early-Delayed Start at week 72	-1.592	0.762	0.0374	-3.090	-0.093
	431	2 mg Early-Delayed Start at week 72	-0.088	0.681	0.8974	-1.426	1.250
PP	405	1 mg Early-Delayed Start at week 72	-1.594	0.770	0.0392	-3.108	-0.079
	429	2 mg Early-Delayed Start at week 72	-0.024	0.679	0.9723	-1.359	1.312

Mixed Model Repeated Measures Model with categorical week in trial using Unstructured Covariance Matrix  
 Separate analysis by each dose level

Note: This table was copied from sponsor’s study report page 133

### Multiple Imputation Sensitivity Analysis for Separate Data Sets

See page 35 of 3.2.2.1 (Study Design and Analysis Plan) for details of the multiple imputation sensitivity analysis method used to produce the results described below. As for the primary analysis for hypothesis #2, larger deteriorations were evident in the 1 mg delayed-start group compared to the 1 mg early-start group in a multiple imputation sensitivity analysis with treatment effects in favor of rasagiline early treatment (see Table 33).

Similar deteriorations from baseline at Week 72 were evident for the 2 mg early and 2 mg delayed-start groups.

**Table 33 Multiple Imputation Sensitivity Analysis for Hypothesis 2 (Sponsor’s Analysis)**

	<b>Estimate</b>	<b>SE</b>	<b>P-Value</b>	<b>Lower 95% CI Limit</b>	<b>Upper 95% CI Limit</b>
<b>Estimate</b>					
<b>1 mg Early-Delayed Start at week 72</b>	<b>-1.832</b>	<b>0.713</b>	<b>0.0105</b>	<b>-3.232</b>	<b>-0.432</b>
<b>2 mg Early-Delayed Start at week 72</b>	<b>-0.435</b>	<b>0.693</b>	<b>0.5308</b>	<b>-1.802</b>	<b>0.931</b>

Mixed Model Repeated Measures Model with categorical week in trial

Separate analysis by each dose level

Note: this table was copied from study report page 134

*Reviewer’s Comments: There seems to be a rule of thumb that 5 imputations is enough to get a reliable result. The more imputations that are used the less dependent the result should be on the random seed and there isn’t any computational impediment to increasing the number of imputations. It appears that this random seed was not prespecified as it should ideally have been, therefore we don’t know how many times the MI procedure was run before the stated result was obtained. It seems a little surprising to this reviewer that the MI result for 1mg has a smaller p-value than for the primary analysis but they are fairly close and this may be due to the larger sample size and the imputation being based on UPDRS assessments throughout the trial rather than just in the ACTIVE phase.*

*The sponsor’s code for the multiple imputation was requested post-submission since it was not submitted with the application and it contains a few potentially important details that were not specified in the analysis plan. In particular,*

- The number of burn-in iterations for the first step of the 3 step multiple imputation process, i.e., the MCMC imputation to get a monotone missing pattern, was not prespecified nor was the number of post burn-in iterations. Both of these can affect the results and affect the reliability of the results, i.e., how similar a result would be obtained if we re-did the entire multiple imputation process with a different random seed.*
- There were separate regression models for each UPDRS timepoint and each such model included all of the previous UPDRS assessments including those from the placebo controlled phase as covariates as well as age, sex, site, and time to need of additional PD treatment. It was not clear from the analysis plan if or how time or earlier assessments were involved in the model.*
- The random seed was not prespecified thus it can’t be ruled out that the particular seed could have been selected after examining results from several seeds.*
- Also, it appears that the model didn’t converge with the unstructured covariance for some of the sponsor’s 5 imputed complete datasets so the reported analysis of the combined multiple imputations was based on a different assumption about the within patient covariance structure than the primary analysis. This may affect the comparison of the multiple imputation result with the primary analysis result.*
- Both phases of the trial were used to impute for the active phase. This is at least somewhat reasonable but it was not clear in the plan and it could bias the imputations in*

*the direction of the placebo controlled phase. If it is reasonable to base the primary analysis for hypothesis 2 on the ACTE dataset then it would also be reasonable and important to impute only the active phase based on only the ACTE dataset. This reviewer found a 1mg delayed-early estimated difference at Week 72 of 1.64 +/- .754, p=0.0302 based on a similar MI approach restricted to the ACTE subset (imputing only week 48 and beyond and not using any post-baseline measurements from the PC phase in the imputation models)*

### **Imputation with the Means of the Delayed-Start Groups Sensitivity**

#### **Analysis for Separate Data Sets**

As for the primary analysis for hypothesis #2, a larger deterioration was evident in the 1 mg delayed-start group compared to the 1 mg early-start group using imputation with delayed-start group means; with treatment effects in favor of rasagiline early treatment (see Table 34 and Table 35). However, the 1MG difference was smaller than in the primary analysis and not significant at the multiplicity adjusted significance level of 0.0250. Similar deteriorations from baseline at Week 72 were evident for the 2 mg early and 2 mg delayed-start groups.

Table 34 Change from Baseline to Week 72 in Total UPDRS Score –  
Missing values are Imputed with Delayed-Start Group Means

	<b>Estimate</b>	<b>SE</b>	<b>P-Value</b>	<b>Lower 95% CI Limit</b>	<b>Upper 95% CI Limit</b>
<b>Treatment Group</b>					
<b>1 mg Delayed Start</b>	<b>3.468</b>	<b>0.419</b>	<b>&lt;.0001</b>	<b>2.644</b>	<b>4.292</b>
<b>1 mg Early Start</b>	<b>2.276</b>	<b>0.426</b>	<b>&lt;.0001</b>	<b>1.440</b>	<b>3.112</b>
<b>2 mg Delayed Start</b>	<b>2.479</b>	<b>0.384</b>	<b>&lt;.0001</b>	<b>1.725</b>	<b>3.233</b>
<b>2 mg Early Start</b>	<b>2.532</b>	<b>0.387</b>	<b>&lt;.0001</b>	<b>1.772</b>	<b>3.291</b>

Mixed Model Repeated Measures Model with categorical week in trial

Separate analysis by each dose level

Note: this table copied from page 135 of sponsor’s study report

Table 35 Comparison of Changes from Baseline to Week 72 in Total UPDRS Scores – Imputed ITT Data Analysis Set after using Imputation with Delayed-Start Group Means

	Estimate	SE	P-Value	Lower 95% CI Limit	Upper 95% CI Limit
<b>Comparison</b>					
<b>1 mg Early-Delayed Start at week 72</b>	<b>-1.192</b>	<b>0.586</b>	<b>0.0422</b>	<b>-2.343</b>	<b>-0.042</b>
<b>2 mg Early-Delayed Start at week 72</b>	<b>0.053</b>	<b>0.530</b>	<b>0.9203</b>	<b>-0.987</b>	<b>1.093</b>

Mixed Model Repeated Measures Model with categorical week in trial  
 Separate analysis by each dose level

Note: this table copied from page 135 of sponsor’s study report

**Reviewer’s Assessment of the Impact of Missing Data due to Deaths**

There was 1 death in 1MG early group in the placebo controlled phase and 1 death in 1MG early group in the active controlled phase and no deaths in any other groups. The 1 Mg early patient that died in the active phase had a change from baseline in UPDRS of +1 at the start of the active phase. This increased to +4 above baseline at week 42 and then went back down to +1 above baseline at week 48, the patient’s last available UPDRS assessment. If we impute the maximum observed UPDRS of 71 in the active phase for the missing week 72 assessment of the patient that died then the results will get worse as follows.

	Estimate	S.E.	p-value
separate dataset 1MG D-E	1.4950	0.7699	0.0528
combined dataset 1MG D-E	1.2409	0.7403	0.0940

Imputing the week 72 change for this death based on the worst observed change (=+35) at week 72 in the 1 mg early or delayed groups and re-doing the combined dataset analysis results in an estimated week 72 difference between 1 MG groups of 1.29, p=.0806. Based on the separate dataset the corresponding result is 1.54, p=0.0434.

Imputing the week 72 change for this death based on the worst observed change (=+29) at week 72 in the 1 mg early or delayed groups among those with a similar baseline score (+/- 2 points) and re-doing the combined dataset analysis results in an estimated week 72 difference between 1 MG groups of 1.31, p=.0736. Based on the separate dataset the corresponding result is 1.57, p=0.0386.

For another general sensitivity analysis for missing data a Diggle Kenward Not Missing at Random (NMAR) model was fit by this reviewer using SAS code obtained from the book *Pharmaceutical Statistics Using SAS: A Practical Guide*. In this Diggle Kenward model the probability of actually obtaining a scheduled UPDRS assessment at a given visit time is assumed to depend on the change from baseline at the last assessment as well as the change from baseline for the given time and these two effects are allowed to vary by treatment group.

The result obtained based on the 1mg separate dataset suggested slightly more informative censoring for the 1mg early group than the 1 mg delayed group and the estimated difference for

1mg delayed-early at week 72 was 1.688,  $p=0.0265 > 0.0250$ . The estimated probability of a non-missing UPDRS assessment at time  $j$  was

$$\text{Prob}(r[j]) = -161.110 + 5.922 * y[j-1] - 9.800 * y[j] \text{ for 1MG Delayed,}$$

$$-161.110 + 9.168 * y[j-1] - 14.007 * y[j] \text{ for 1MG Early}$$

where  $y[j]$  denotes the change from baseline in UPDRS at time  $j$ ,  $j=48, 54, 60, 66, \text{ or } 72$ .

The descriptive statistics for mean change from baseline in UPDRS for the non-ACTE cohort suggest a different pattern than for the ACTE cohort (Table 36). Non-ACTE cohort patients may have transferred to the ACTIVE phase before Week 24. The sponsor pre-specified excluding these patients from the analysis. Although it was prespecified this exclusion seems debatable.

**Table 36 Comparison of Mean UPDRS Change from Baseline between Active Phase Analysis Eligible and Ineligible Patients**

	Cohort											
	Non- ACTE						ACTE					
	Treatment Group						Treatment Group					
	1mg delayed			1mg early			1mg delayed			1 mg early		
	UPDRS Total (Change)			UPDRS Total (Change)			UPDRS Total (Change)			UPDRS Total (Change)		
	N	Mean	Std	N	Mean	Std	N	Mean	Std	N	Mean	Std
<b>Week</b>												
<b>12</b>	32	2.2	7.07	22	3.9	8.05	238	0.4	4.41	251	-1.7	4.99
<b>24</b>	24	6.6	7.37	18	4.0	5.79	237	2.1	5.77	251	-1.6	5.33
<b>36</b>	10	12.5	7.37	9	4.2	4.60	226	3.3	6.82	246	0.3	5.57
<b>42</b>	44	8.6	10.90	29	5.2	10.73	236	1.8	6.64	251	-0.1	6.31
<b>48</b>	20	3.4	8.17	10	1.9	7.17	245	2.1	6.79	254	0.7	7.14
<b>54</b>	16	-0.3	5.01	10	5.1	7.08	231	2.2	7.34	245	0.6	7.03
<b>60</b>	13	-4.3	6.45	8	2.1	8.98	226	2.3	7.44	248	1.4	7.43
<b>66</b>	14	-1.6	8.33	8	1.6	6.85	221	2.8	8.00	237	1.8	7.89
<b>72</b>	14	-1.3	7.27	8	3.2	8.31	219	3.4	8.98	231	1.9	8.11

An exploratory sensitivity analysis by this reviewer based on the hypothesis 2 analysis method but also including non-ACTE patients in the analysis of the active phase for active phase assessments beyond Week 42 yielded an estimated 1 mg delayed-early difference at Week 72 of 1.47,  $p=0.0444$ . The number of patients in this separate 1 mg analysis was 515.

### Hypothesis 3

Since linearity was not rejected by the sponsor's nonlinearity test ( $p=0.0893>.05$ ), the alternative model to handle the nonlinear case was not required in the sponsor's opinion. As described for Hypothesis #2), analyses were conducted on separate data sets for 1 mg and 2 mg for the primary analysis of Hypothesis #3. This reviewer notes however, that the nonlinearity test was based on the combined dataset and there are other indications of significant nonlinearity to be described below after the presentation of the sponsor's results for hypothesis 3.

Results from repeated measures analysis providing the changes in UPDRS per week (slope estimates) for each of the 4 treatment groups during the active treatment phase are displayed in Table 37.

Table 37 Change Per Week (Slope) in Total UPDRS Score during Active-Treatment Phase - ACTE Data Analysis Set, Separate Data Sets

		Estimate	SE	P-Value	Lower 95% CI Limit	Upper 95% CI Limit
Data Analysis Set	Estimate					
ACTE	1 mg Delayed Slope	0.085	0.016	<.0001	0.054	0.116
	1 mg Early Slope	0.085	0.015	<.0001	0.055	0.115
	2 mg Delayed Slope	0.065	0.015	<.0001	0.037	0.094
	2 mg Early Slope	0.094	0.014	<.0001	0.066	0.122

Repeated Measures Mixed Linear Model with Random Intercept and Slope using Unstructured Covariance Matrix between Intercept and Slope Estimates

Separate analysis by each dose level

Note: This table was copied from page 141 of sponsor's study report

The treatment effects (the differences between each of the early-start and delayed-start treatment groups in the changes in UPDRS per week during the active-treatment phase) are displayed in Table 38. An identical rate of deterioration in UPDRS was evident in the 1 mg delayed-start and 1 mg early-start treatment groups (each 0.085 UPDRS units/week). The point estimate for the difference between the treatment groups in the change in UPDRS per week was 0.00 with 90% confidence intervals of -0.036 to 0.036. Since the upper confidence interval clearly did not exceed the predefined upper non-inferiority boundary of 0.15 UPDRS units/week, the null hypothesis was rejected and the 1 mg early start was declared as not inferior to the delayed-start, with regard to deterioration rates in the active phase (see Table 38).

**Table 38 Comparison of Changes per Week (Slopes) in Total UPDRS (Separate Datasets)**

			Estimate	SE	Lower 90% CI Limit	Upper 90% CI Limit
Data Analysis Set	No. of Subjects	Comparison				
ACTE	489	1 mg Early-Delayed Start Slope Difference	0.000	0.022	-0.036	0.036
	507	2 mg Early-Delayed Start Slope Difference	0.029	0.020	-0.005	0.062

Repeated Measures Mixed Linear Model with Random Intercept and Slope using Unstructured Covariance Matrix between Intercept and Slope Estimates

Separate analysis by each dose level

Note: This table was copied from page 142 of sponsor's study report

The results for the combined dataset were very similar (Table 39).

**Table 39 Comparison of Group Differences in Changes per Week (Slopes) in Total UPDRS (Combined Dataset)**

	Estimate	SE	Lower 90% CI Limit	Upper 90% CI Limit
<b>Comparison</b>				
<b>1 mg Early-Delayed Start Slope Difference</b>	-0.001	0.017	-0.029	0.027
<b>2 mg Early-Delayed Start Slope Difference</b>	0.030	0.017	0.003	0.058

Random Intercept and Slope Model using Variance Components (VC) Covariance Matrix between Intercept and Slope Estimates assuming Linearity

Note: This table was copied from page 142 of sponsor's study report

The sponsor asserts that linearity of the change from baseline in UPDRS over weeks 48-72 was not rejected (Likelihood Ratio Test  $p=0.0893$ ). This result was for the combined dataset but they used the separate datasets for hypothesis 3 (parallelism). If one is using the separate datasets then it seems to this reviewer that one should perform the linearity test on the separate datasets as well. Also, since the 2mg early vs. delayed comparison was not positive for hypothesis 2 (mean difference at week 72) hypothesis 3 (parallelism) for 2mg early vs. delayed doesn't serve the intended purpose (to suggest that a difference at week 72 would persist). In fact, based on the model for hypothesis 2 the estimated difference between 2 mg delayed and 2 mg early is not significant at week 48 (.4495,  $p=.3770$ ) and numerically in the wrong direction at week 72:  $-.3564$   $p=.6028$ . Thus, the question of linearity for the 2MG groups is not very relevant. If this separate linearity test is done for 1MG then the p-value is .0435, which suggests that linearity does not hold across the 1MG Early and 1MG delayed groups. In the same vein, an F-test for any significant quadratic terms (i.e, for either 1MG early or 1MG delayed group) based on a repeated measures model with unstructured within patient covariance and linear and quadratic Week terms for each group, as well as site and baseline covariate effects, gives a p-value of 0.0012. Furthermore, the prespecified test for linearity doesn't even involve the same model which is the basis for testing hypothesis 3. In particular, the linearity test uses a repeated statement to model the within patient covariance structure, i.e., it assumes a very general correlation pattern between the random deviations from the model within a patient. On the other hand, the model for hypothesis 3 assumes that a subject's random deviations from the model are explained by a random subject specific intercept and slope and that after accounting for a

subject's random intercept and slope any deviations from the model within a patient are uncorrelated.

There was initially a convergence problem in SAS when the population (non-random coefficient) slope model involved in the sponsor's test for non-linearity was fit with unstructured within patient correlation to the 1 MG separate dataset. However, when the covariance parameter estimates obtained from the hypothesis 2 primary analysis model fit were supplied as starting values for the optimization routine the slope model converged. This permitted the sponsor's likelihood ratio test for nonlinearity with unstructured within patient correlation to be obtained for the 1MG separate dataset. The resulting chi-square likelihood ratio test statistic was  $-13121.6 + 13134.5 = 12.9$  with 6 degrees of freedom, which corresponds to a p-value of  $p=0.0435$ . Thus, there was an indication of significant non-linearity over weeks 48-72 in the 1MG separate dataset.

The same result for the likelihood ratio test between the models involved in the non-linearity test with the unstructured covariance within patient was reproduced with a different software package, STATA, and there were no convergence problems using STATA. Therefore, this reviewer concludes that the 1MG UPDRS change from baseline data was significantly non-linear over weeks 48-72 meaning that a constant slope is not adequate to describe the data over this period. Thus, it is not clear that the observed 1MG delayed – early difference was approximately constant over weeks 48-72 and it is, therefore, less clear that the observed difference at week 72, such as it is, would persist.

Figure 7 illustrates the non-linearity of the 1MG change in UPDRS data over the active phase. It shows linear and quadratic fits as well as the group least squares means for each visit based on the MMRM model (hypothesis 2) of the 1MG separate dataset. The upper curve is the delayed group that hasn't had the benefit of treatment in the placebo controlled phase and therefore starts higher. One can observe from the figure that for each group the deviations of the visit specific means from the corresponding quadratic fit are smaller than those from the linear fit, particularly at active phase week 12 (overall trial week 48) and active phase week 36 (overall week 72). The graph also shows that the difference in LSMeans at week 72 is about 50% smaller than the difference at week 36. This reviewer found based on an MMRM model that as a percentage of the difference at week 36 between 1mg early and 1 mg delayed the difference at Week 72 is 53% with a 95% C.I. of 11% to 95%.

This reviewer found based on the model for hypothesis 2 that the difference between the 1mg treatment effect averaged over the visits at weeks 48 and 54 and the corresponding 1mg effect averaged over the visits at weeks 66 and 72 is estimated to be  $-0.08656 \pm 0.4056$  S.E. with a 95% confidence interval of  $-0.8836$  to  $0.7105$ . The 90% confidence interval's lower limit is  $-0.7550$ . The negative sign suggests that at least numerically some effect was lost in going from the week 48 and 54 period to the week 66 and 72 period. This was based on the 1mg separate data analysis but the result for the combined dataset was very similar.

The standard error of this difference of effect by period is .4056. If the lower limit was -2.7 so that the difference was non-inferior but on the boundary of inferiority the upper limit of the 95% confidence interval would be about  $4 \times .4056$  points higher, that is  $-2.7 + 4 \times .4056 = -1.6224$  (the 90% confidence interval would be  $-2.7 + 2 \times 1.644 \times .4056 = -1.37$ ). The estimated difference between 1 mg delayed and 1 mg early groups averaged over the two visits week 48 and week 54 was  $1.61 \pm .535$ , which has an upper 95% limit of 2.66 and an upper 90% limit of 2.49. This non-inferiority margin would have allowed one to conclude non-inferiority even if all of the effect averaged over the visits at week 48 and 54 was lost. Therefore, this margin seems far too liberal.

Figure 7 Change from Baseline in UPDRS during Active Phase (ACTE) for 1MG

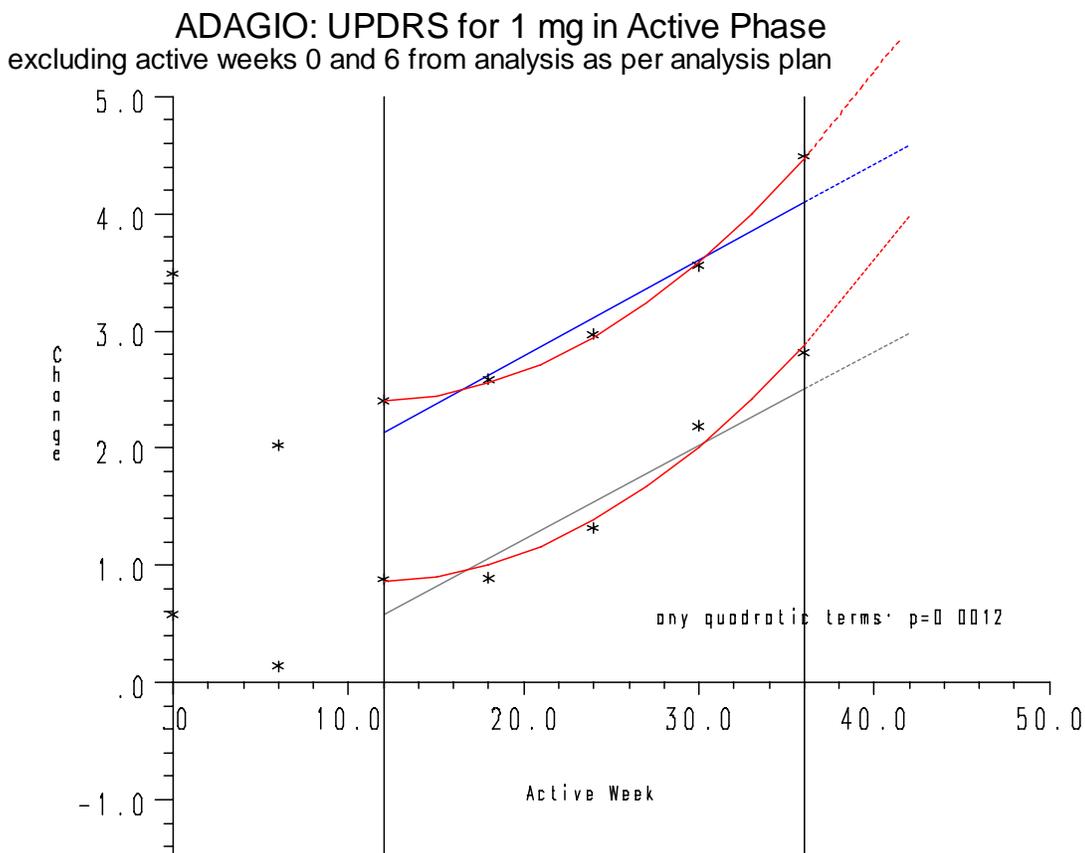
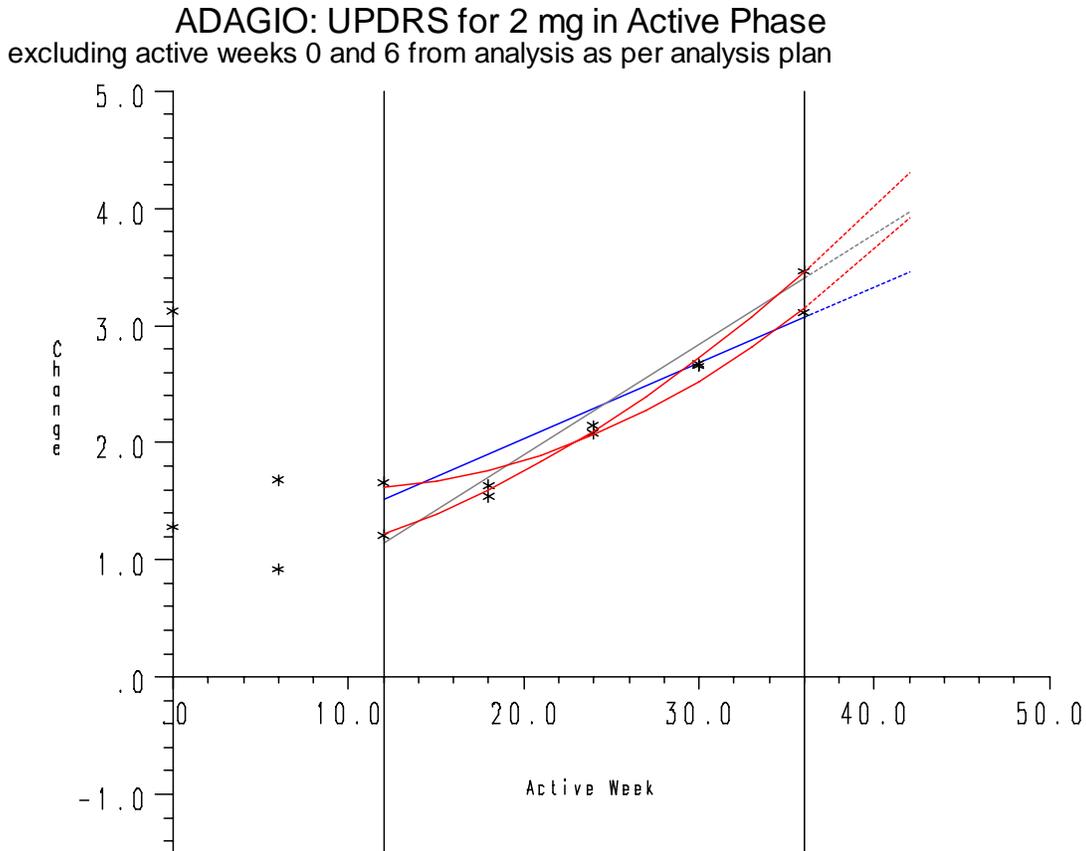


Figure 8 shows the corresponding figure for the 2MG separate dataset.

Although the two groups' lines and curves are all relatively close the early and delayed lines as well as quadratic curves cross indicating that although the delayed group started off worse they ended up numerically better than the early group.

**Figure 8 Change from Baseline in UPDRS during Active Phase (ACTE) for 2MG**



One can observe from the graph for 1MG in Figure 7 that the linear fit does not fit the Week 72 group means very well. Also, the quadratic fits deviations from the visit means are as small or smaller than those of the linear fits.

There was a convergence problem when a random quadratic term for subject was added to the hypothesis 3 analysis model for the separate 1MG dataset as another check for nonlinearity. This may be caused by the differing scales of the various parameters, e.g., at week 72 the quadratic term coefficient is multiplied by  $72 \cdot 72 = 5184$  as compared to multiplication by 1 for the intercept and by 72 for the slope parameter so some parameters may be much larger than others and this can cause instability in the model fit optimization process. One method for dealing with this is standardizing the slope and quadratic terms in the model by replacing week with  $[\text{week} - \text{avg}(\text{week})] / \text{stddev}(\text{week})$  and  $\text{week}^2$  with  $[\text{week}^2 - \text{avg}(\text{week}^2)] / \text{stddev}(\text{week}^2)$ . The resulting quadratic model using standardized linear and quadratic terms with random intercept, slope and quadratic terms for each subject converged and an F test for any significant quadratic terms based on this model had a p-value of  $p = .0026$ . This is more evidence of non-linearity during weeks 48-72 of the active phase for the 1MG early and delayed groups.

This model suggests that the slope difference (1MGDelayed- 1MG Early) decreases slightly with increasing week.  $-2*.1126*W/1066 +.1373/9.03 = -.00021*W+.0152$ . The slope difference is 0 at week 72 and would become negative, which corresponds to the 1MG early start slope being bigger, beyond that.

If the UPDRS change relationship with time is nonlinear we really can't say what happens beyond the range of the data but it seems to this reviewer that we can't have as much confidence that the observed separation of the 1MG groups at Week 72 will be maintained.

The prespecified non-inferiority margin would allow the early group to have a slope as much as .15 bigger and still be considered non-inferior. If the estimated lines had a difference of .15, this continued to hold beyond week 72 and the group difference is 1.6 at week 72 then with such a slope difference the group difference would decrease to 0 at week 82.67. In order to not be significantly inferior the slope difference would have to be slightly smaller than 0.15 (the observed standard error was 0.022) but the above result would only increase by 2 or 3 weeks if we made this adjustment.

It seems a little questionable whether a slope difference that could lead to a mean group difference in UPDRS change from baseline of 0 in just over 10 weeks beyond the end of the study is enough. This would seem to imply that the non-inferiority margin is too liberal and/or the clinical meaningfulness of the observed 1MG group difference at week 72 is questionable.

### 3.3 Evaluation of Safety

Safety is not reviewed in this document. Please see the medical review.

## 4. FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

### 4.1 Gender, Race, Age, and Geographic Region

#### 4.1.1 Gender

In ADAGIO approximately 60% of subjects in all 4 treatment groups were male.

**Table 40 ADAGIO: Week 36 Estimated Differences in UPDRS Change from Baseline by Gender**

	GROUP	DIFF FROM PLACEBO	STD. ERR.	P-VALUE*
Female	1MG	3.5685	0.7920	<.0001
Male	1MG	2.6784	0.6284	<.0001
Female	2MG	3.6607	0.7766	<.0001
Male	2MG	2.8520	0.6314	<.0001

\*Note: Week 36 estimated diffs based on MMRM (not random slope model because of non-linearity)

Table 41 shows the Week 72 estimated differences in UPDRS change from baseline between 1mg delayed and 1mg delayed groups.

**Table 41 ADAGIO: UPDRS change from baseline between 1mg groups by GENDER**

GENDER	1MG DIFFERENCE DELAYED -EARLY	STD. ERR.	P-VALUE
Femal e	4.2387	1.2084	0.0005
Mal e	0.02762	0.9557	0.9770

An F test for a significant interaction between the effect of gender and treatment group on UPDRS Change was nominally significant,  $p=0.0434$ . Similarly, the week 72 difference between gender specific (male-female) 1MG treatment differences (delayed-early) is nominally significant:  
estimate= -4.2111,  $p=0.0074$ .

Females accounted for only (1MG Delayed N=91 1MG Early N=98 ) 39% of the 1MG portion of the Active Phase Analysis Set as compared to males (147 1MG Delayed and 153 1MG Early). Note that there was a nominally significant difference in mean baseline UPDRS score within 1MG females: 1 MG Delayed Mean= 17.87 vs. 1 MG Early Mean=20.78 ,  $p=0.0143$ . This might explain part of the observed treatment group difference in this subgroup. There was no apparent corresponding baseline difference in Males.

**Figure 9 ADAGIO: Gender Specific Patterns of LS Mean UPDRS Changes by Visit for 1MG groups**

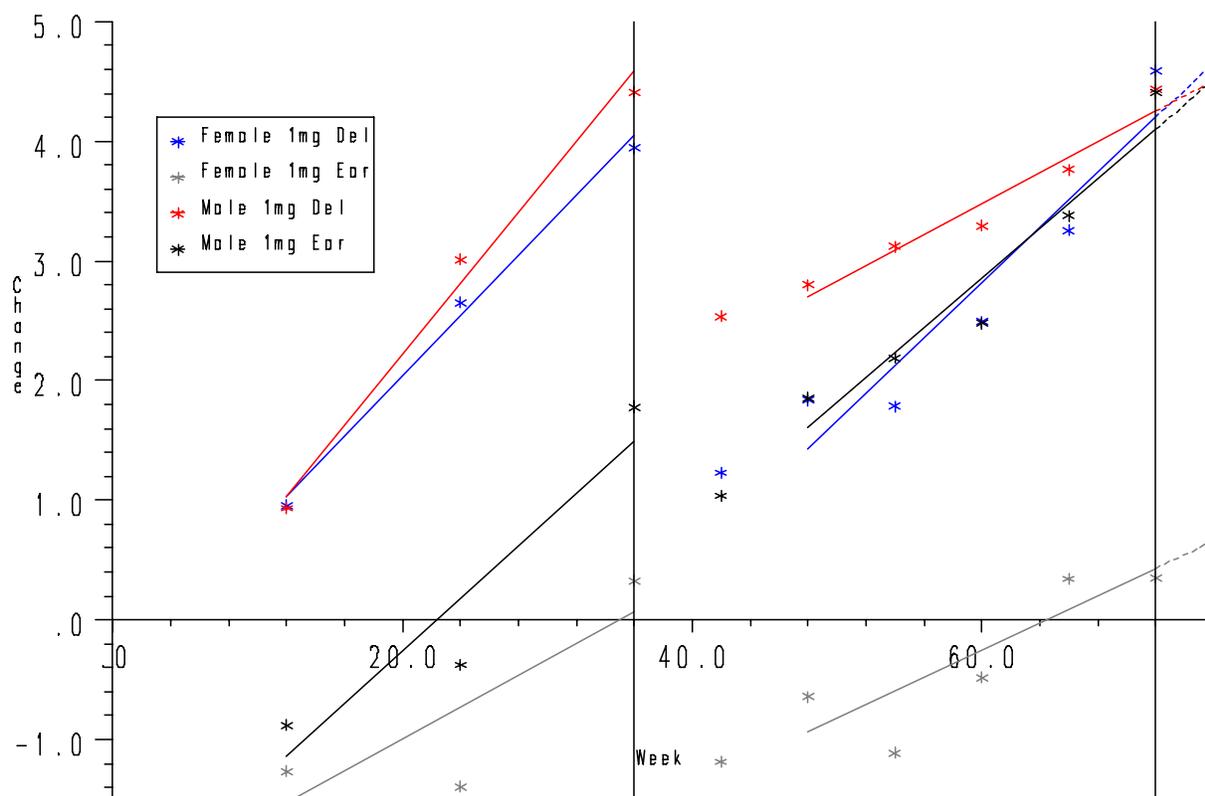


Table 42 shows the estimated slope differences (1mg delayed-1mg early) over Weeks 48 to 72 by Gender. The slope difference in Males numerically favors the 1mg delayed group over the 1 mg early group.

**Table 42 ADAGIO: 1MG Slope Differences Delayed-Early over Weeks 48 to 72 by Gender**

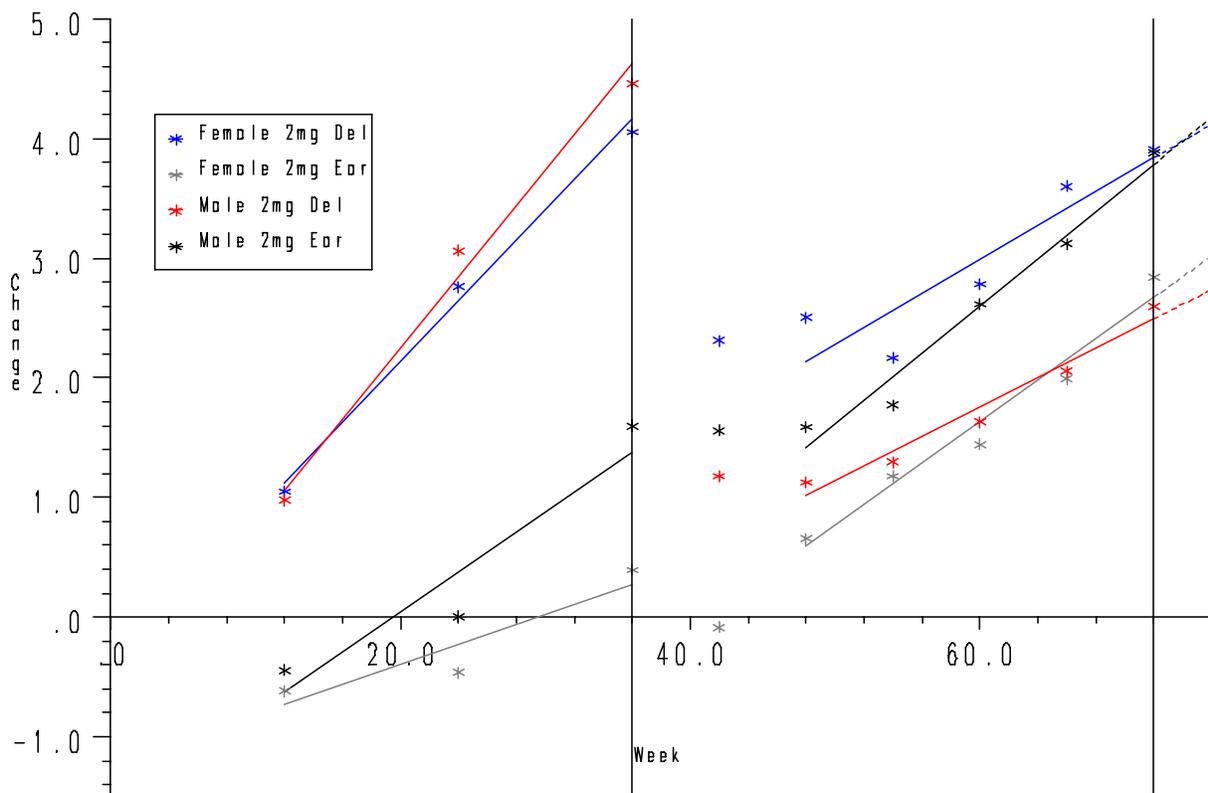
GROUP	ESTIMATED DIFF	STD.ERR.	P-VALUE	95% LOWER	95% UPPER
Female	0.05860	0.03413	0.0867	-0.0085	0.1257
Male	-0.03920	0.02780	0.1592	-0.0938	0.01543
Diff Male-Female	-0.09780	0.04401	0.0268	-0.1843	-0.0113

Although neither of two gender specific slope differences (1MG Del – 1MG Ear) is significantly different from 0 they have different signs and the difference between the female slope difference

(1mg delayed-early) and the male slope difference (1mg delayed-early) is nominally significant,  $p=0.0268$ .

Figure 10 shows Gender Specific Patterns of LS Mean UPDRS Changes by Visit for the 2MG groups. The 2mg delayed –early difference was roughly the same size in absolute value across gender but had a different sign (favoring early in Females and delayed in Males). In Males 2mg Early was numerically worse than 2mg Delayed from week 48 onwards and the slope of 2 mg early was also numerically bigger. However, none of these differences were nominally significant.

Figure 10 ADAGIO: Gender Specific Patterns of LS Mean UPDRS Changes by Visit for 2MG groups



In TEMPO 36% of randomized patients were female. The 2mg delayed – 2mg early difference was 2 points bigger in Males than in Females, however the difference did not reach statistical significance.

**Table 43 TEMPO: Week 52 Group Differences in UPDR Change by Gender**

GROUP	ESTIMATED DIFFERENCE FROM PLACEBO/DELAYED	STD.ERR.	P-VALUE
Female 2MG Early	-0.6292	1.7297	0.7163
Male 2MG Early	-2.8086	1.3194	0.0342
Female-Male 2MG Early	2.1795	2.1963	0.3219
Female 1MG Early	-0.6763	1.8024	0.7078
Male 1MG Early	-0.8488	1.2396	0.4941

Note: These estimates were based on an MMRM model for weeks 32-52

In summary, there was significant inconsistency in treatment differences at the end of the active phase by Gender in ADAGIO for 1MG with the suggestion of an effect in Females but very little if any effect apparent in Males, the larger subgroup (60%). In TEMPO, the treatment group difference between 2mg delayed and 2mg early was larger in Males than Females but not significantly so. Overall, it seems there is no consistent treatment effect by Gender.

#### **4.1.2 Race**

The great majority of subjects in ADAGIO were Caucasian (97.7%). In TEMPO also, 94.8% were Caucasian. Therefore, it is not possible to reliably determine whether there are any race differences by treatment based on the available data.

### 4.1.3 Age

The mean age of subjects in the ADAGIO study was approximately 62 years in all treatment groups with a majority in the 55 to 65 (36%) and 65 to 75 (34%) age categories. In the Active Phase Analysis set 42% were over age 65.

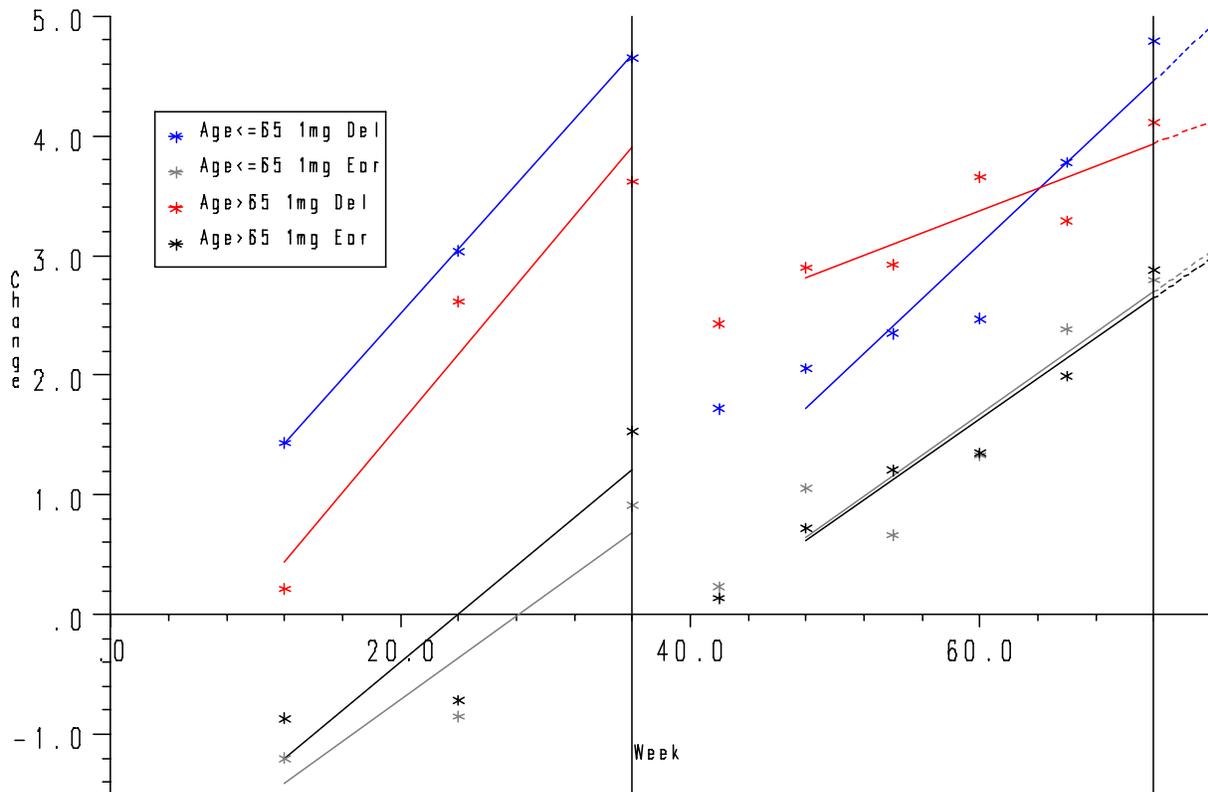
There was no indication of treatment differences between the age groups in the UPDRS change from baseline results during the placebo controlled phase.

**Table 44 ADAGIO: LS Mean differences in UPDRS Change at Week 36 by Age Group**

AGE SUBGROUP/TRT GROUP	ESTIMATED DIFFERENCE FROM PLACEBO	STD.ERR.	P-VALUE
≤65 Week36 1MG Early	3.6983	0.6550	<.0001
65+ Week36 1MG Early	2.1608	0.7459	0.0038
≤65 Week36 2MG Early	3.4547	0.6473	<.0001
65+ Week36 2MG Early	2.8671	0.7606	0.0002

Figure 11 shows the pattern of UPDRS changes from baseline over time by Age Group for the 1mg early and delayed groups.

Figure 11 ADAGIO: Pattern of LS Means Changes from Baseline in UPDRS by Visit for 1MG by Age Group



The slope differences (between 1mg delayed and 1 mg early) over weeks 48 through 72 had opposite signs between the age groups but the difference (treatment by gender interaction) was not significant (Table 45).

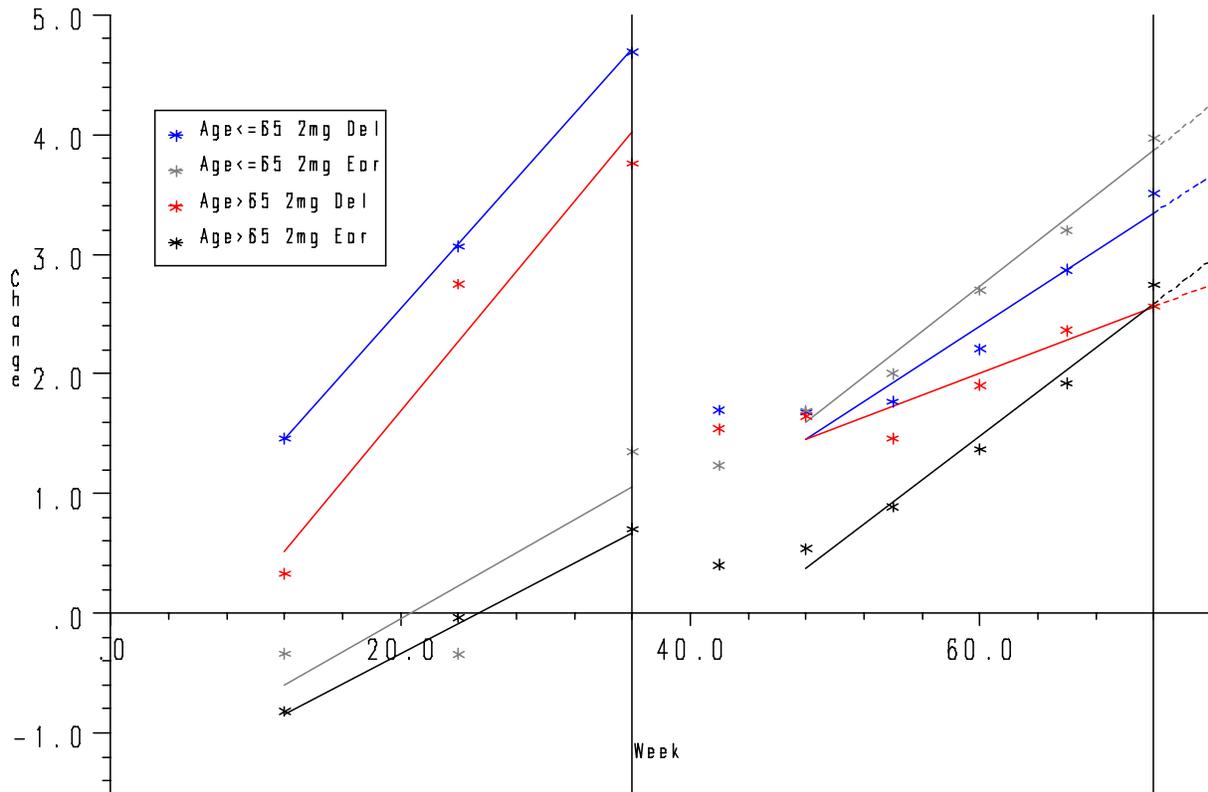
**Table 45 ADAGIO: Age Group specific Slope Differences over Weeks 48-72 for 1MG delayed-early**

GROUP	SLOPE DIFFERENCE 1MG DELAYED-EARLY	STD.ERR.	P-VALUE	LOWER 95%	UPPER 95%
Age≤65	0.02820	0.02880	0.3281	-0.02841	0.08481
Age>65	-0.03768	0.03252	0.2472	-0.1016	0.02623
AgeGrp Diff	-0.06588	0.04344	0.1301	-0.1513	0.01949

Overall, there was no difference at week 72 for 2 mg early vs. delayed as reported in the hypothesis 2 analysis. Using the same approach on age subgroups, we find that in both age

subgroups 2 mg Early was numerically worse than 2 mg delayed at week 72. Figure 12 shows the pattern of UPDRS change from baseline results by age group for the 2mg groups.

**Figure 12 ADAGIO Change from Baseline in UPDRS Profiles for 2MG by Age Group**



In TEMPO the average age was 60.8 and ages ranged from 32 to 92. Forty three percent of the active phase efficacy dataset patients were age 65 or above.

**Table 46 TEMPO: Age Group Specific LS Group Mean Differences in UPDRS Change at Week 52**

GROUP	ESTIMATED DIFFERENCE FROM 2MG DELAYED	STD.ERR.	P-VALUE
≤65 2MG E	-1.7405	1.5537	0.2636
>65 2MG E	-2.4233	1.4089	0.0865
≤65 - >65 Diff. 2MG E	0.6828	2.1120	0.7467
≤65 1MG E	0.4416	1.5807	0.7802
>65 1MG E	-1.7975	1.3406	0.1811

#### 4.1.4 U.S. vs. Non-U.S.

Thirty two percent of the ITT subjects in the ADAGIO study were from U.S. sites. Table 47 shows the distribution of subjects by Country.

Table 47 ADAGIO Distribution of ITT Subjects by Country

TVP1012/500 (ADAGIO) - ITT	1 mg Delayed Start		1 mg Early Start		2 mg Delayed Start		2 mg Early Start		All	
	N	%	N	%	N	%	N	%	N	%
<b>All</b>	<b>298</b>	<b>100.0</b>	<b>288</b>	<b>100.0</b>	<b>295</b>	<b>100.0</b>	<b>293</b>	<b>100.0</b>	<b>1174</b>	<b>100.0</b>
<b>Argentina</b>	17	5.7	17	5.9	18	6.1	16	5.5	68	5.8
<b>Austria</b>	2	0.7	2	0.7	2	0.7	2	0.7	8	0.7
<b>Canada</b>	28	9.4	27	9.4	25	8.5	29	9.9	109	9.3
<b>France</b>	11	3.7	11	3.8	14	4.7	12	4.1	48	4.1
<b>Germany</b>	27	9.1	28	9.7	28	9.5	26	8.9	109	9.3
<b>Hungary</b>	8	2.7	8	2.8	7	2.4	6	2.0	29	2.5
<b>Israel</b>	22	7.4	23	8.0	21	7.1	21	7.2	87	7.4
<b>Italy</b>	24	8.1	25	8.7	25	8.5	28	9.6	102	8.7
<b>Netherlands</b>	9	3.0	9	3.1	9	3.1	9	3.1	36	3.1
<b>Portugal</b>	4	1.3	4	1.4	4	1.4	4	1.4	16	1.4
<b>Romania</b>	25	8.4	24	8.3	24	8.1	25	8.5	98	8.3
<b>Spain</b>	12	4.0	12	4.2	12	4.1	12	4.1	48	4.1
<b>United Kingdom</b>	7	2.3	7	2.4	8	2.7	9	3.1	31	2.6
<b>USA</b>	<b>102</b>	<b>34.2</b>	<b>91</b>	<b>31.6</b>	<b>98</b>	<b>33.2</b>	<b>94</b>	<b>32.1</b>	<b>385</b>	<b>32.8</b>

Note: Copied from page 116 of sponsor's study report

If we conduct the primary analysis method on the US subgroup only, leaving out the other non-US data and using the separate 1MG dataset, we find a 1 MG Early vs. Delayed week 72 difference in LS Mean changes from baseline in UPDRS Total of 2.02,  $p=.1504$  (N=157 or 32% of 489). Recall that the Overall result was 1.68,  $p=.0250$  N=489. The corresponding result for the larger Non-US subgroup is 1.4579 +/- 0.8852 S.E.,  $p=0.1006$  (N=332).

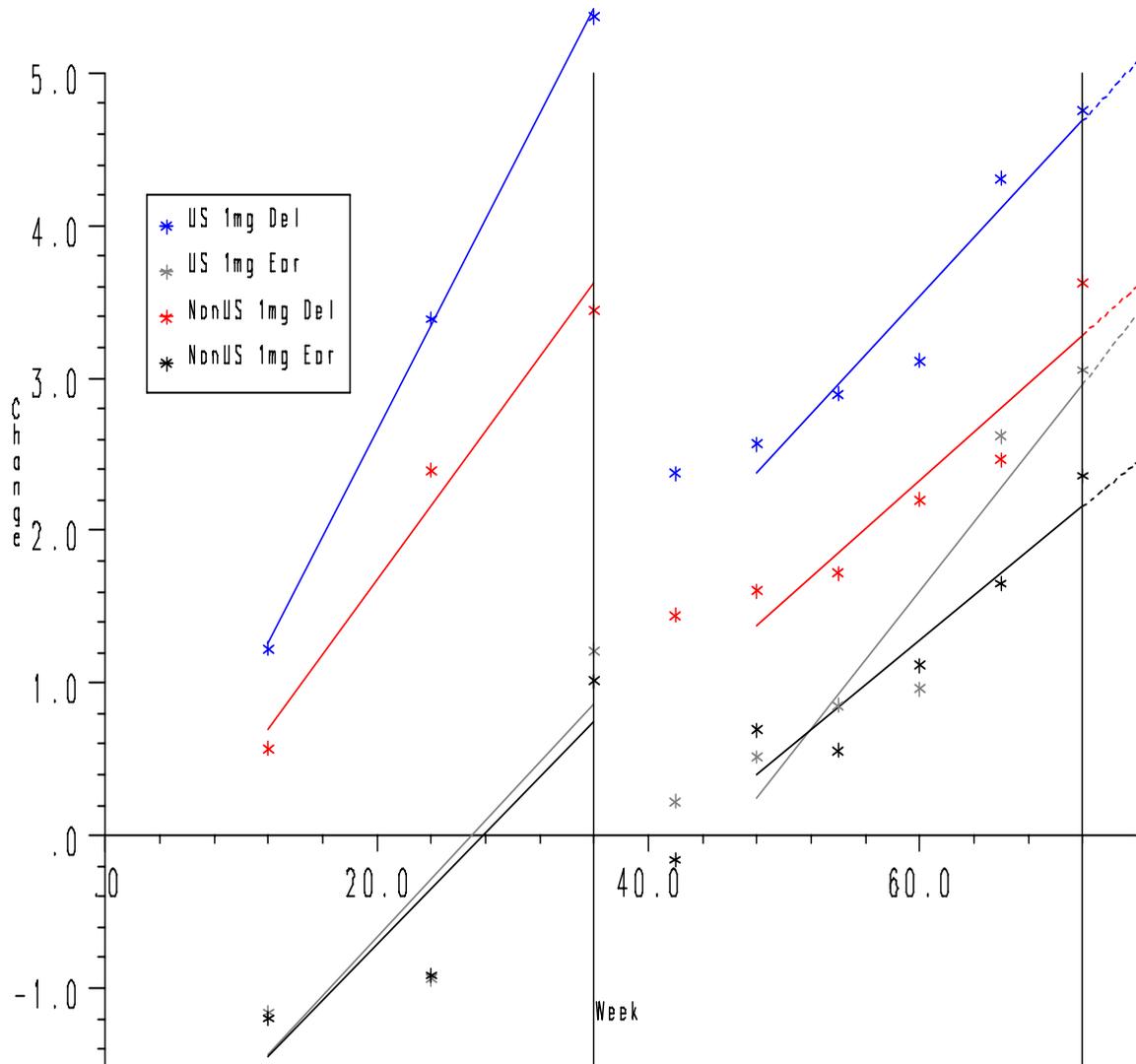
If we adjust for US /Non-US (combine all US sites and, separately, all non-US sites) instead of each individual site then the estimates for 1MG Early vs. 1MG Delayed at Week 72 are considerably different:

US	1.7059	+/-	1.4247	$p=.232$
Non-US	1.2687	+/-	0.9774	$p=.195$

Despite the numerical difference there was no compelling evidence of a differential treatment effect (interaction) between the US and non-US (interaction test  $p=0.800$ ).

Figure 13 shows the LS Mean changes from baseline in UPDRS over visit by US or non-US subgroup. The pattern of means over time for the 1MG early group in the U.S. in the active phase looks particularly non-linear as do the patterns for some of the other groups.

Figure 13 ADAGIO: Change in UPDRS Mean Pattern US vs. Non-US 1MG Early and 1MG Delayed



## **TEMPO**

About 88% (357/404 ITT subjects) of the study was done in U.S. and 12% was done in Canada. Repeated measures analysis using country instead of site yielded the following estimates: Week 52 difference 2mg delayed -2mg early in US was 1.70 +/- 1.27, p=0.1818 and in Canada it was 3.92 +/- 3.43, p=.2536. For 1MG vs. placebo it was 2.48 +/- 3.33, p=.4565 in Canada and .418 +/- 1.25, p=.7387 in the US. There was not compelling evidence that the treatment effects varied significantly by country, p=0.7884 (test for treatment by country interaction).

## **4.2 Other Special/Subgroup Populations**

### **4.2.1 Baseline UPDRS Score**

In ADAGIO in the ACTE dataset the mean baseline total UPDRS score was 19.8 and the quartiles of Baseline UPDRS score were <14, 14-18.5, 18.5-24, and > 24. The range was 3 to 53. In the 4<sup>th</sup> quartile of the baseline UPDRS score the 1MG early group pattern is really nonlinear in the 2<sup>nd</sup> phase (Figure 15).

Figure 14 ADAGIO Change from Baseline in UPDRS Total By Phase and Baseline UPDRS Total quartile

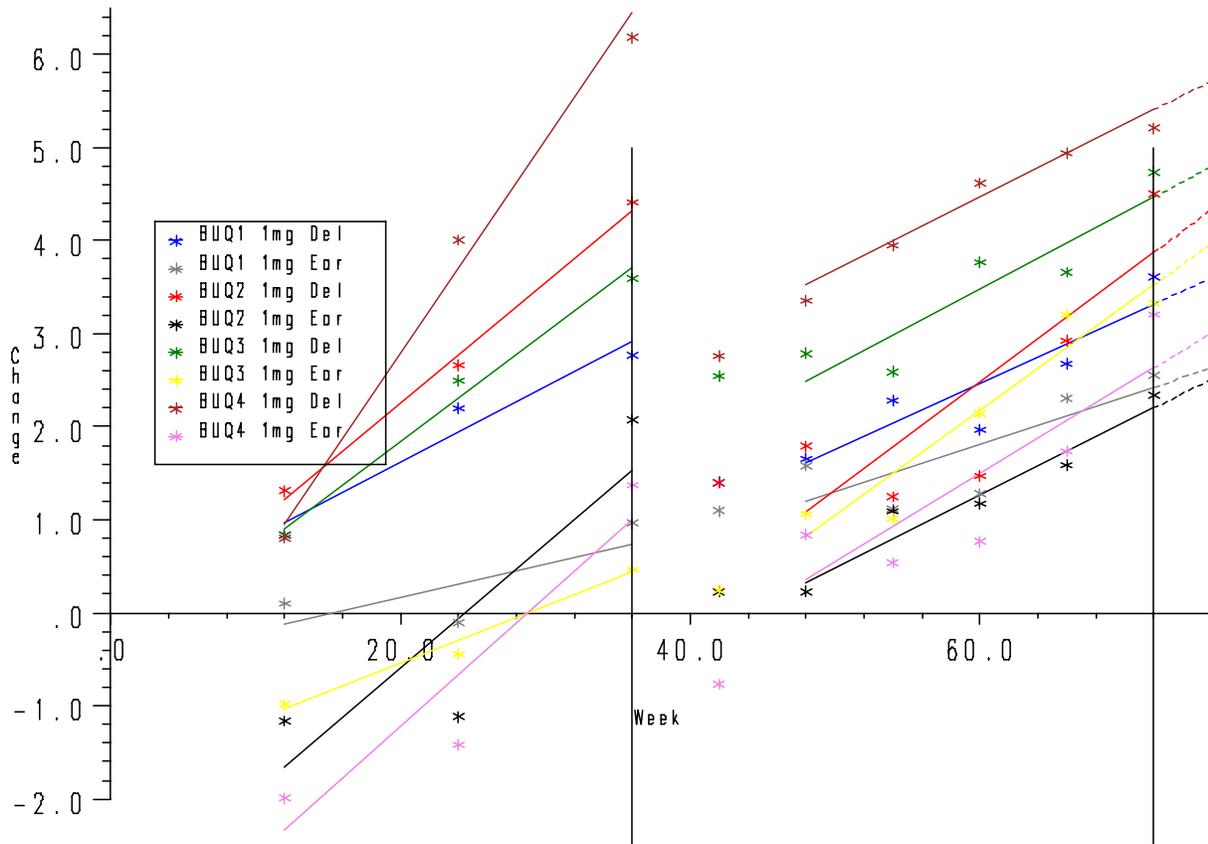
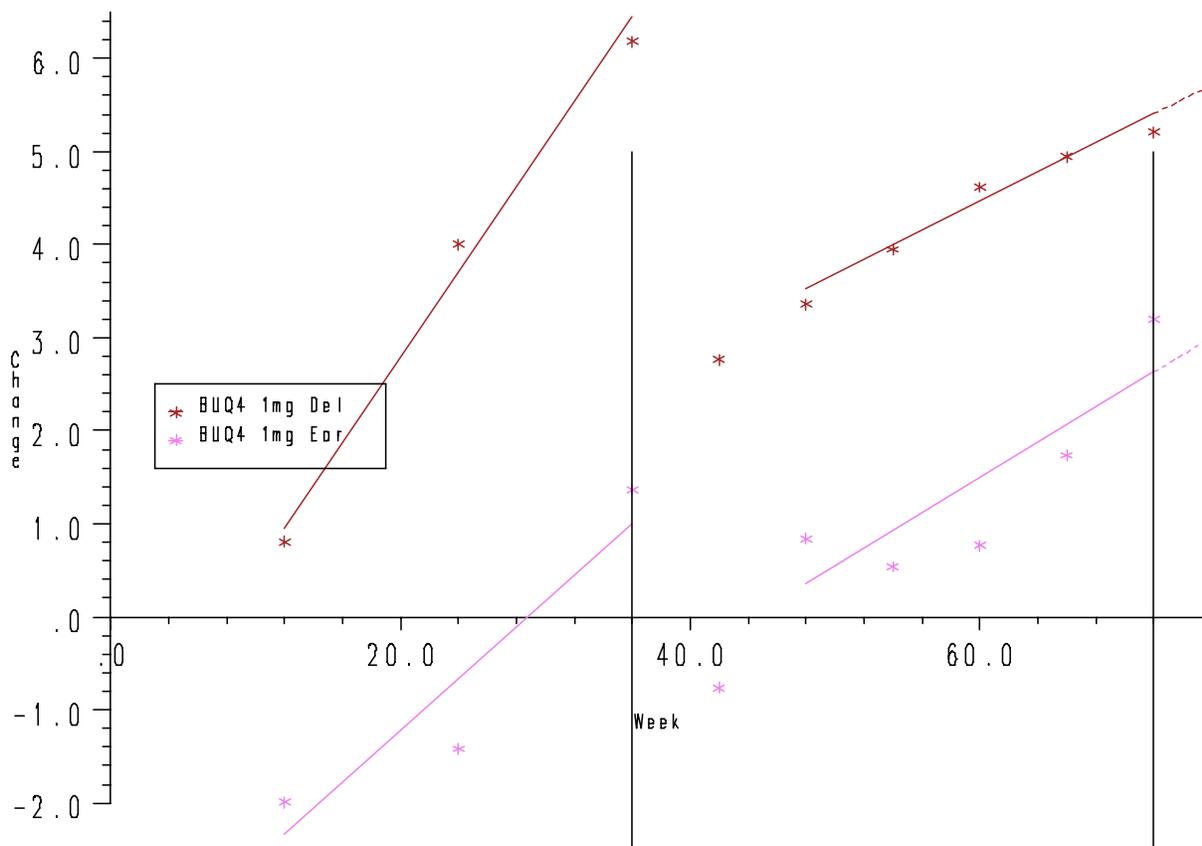


Figure 15 shows just the highest baseline quartile subset. Although, this quartile had the biggest group difference it appears that the pattern of the 1MG Early group is nonlinear and the group difference (1MG delayed – early) appears to be decreasing over time based on the pattern of the visit LS Means.

Figure 15 ADAGIO Change from Baseline in UPDRS Total By Phase in Highest Baseline UPDRS quartile



A test for a linear trend among the week 72 1MG treatment group differences by baseline quartile (i.e., for an increase with baseline quartile) yields an estimated slope of  $-0.8047 \pm 1.75$  S.E.,  $p=0.6457$ . Comparing the 4<sup>th</sup> quartile treatment group difference to the average of the first 3 quartiles we find the difference is not nominally significant either,  $p=0.7980$ . Therefore, there doesn't seem to be compelling evidence that the treatment effect, such as it is, increases with increasing baseline score.

Table 48 Week 72 1MG Treatment Group Differences by Baseline UPDRS Quartile

QUARTILE/OTHER	ESTIMATE	STD. ERR.	P-VALUE
Base Q1 di ff	1.0572	1.4823	0.4761
Base Q2 di ff	2.1683	1.6137	0.1797
Base Q3 di ff	1.4120	1.5566	0.3648
Base Q4 di ff	2.0055	1.5347	0.1920
Test for trend	-0.8047	1.7491	0.6457
Q4 vs. avg of Q1 to Q3	0.4597	1.7953	0.7980

For the 2 MG group difference the linear trend is also not significant,  $p=.6622$ . Notice that the first 3 baseline UPDRS quartile subgroups differences (delayed-early) all have a sign favoring the 2mg delayed group, whereas the 4<sup>th</sup> numerically favors the 2mg early group. Comparing the 4<sup>th</sup> quartile treatment group difference to the average of the first 3 quartiles we find the difference is not nominally significant,  $p=0.0580$ .

**Table 49 Week 72 2MG Treatment Group Differences by Baseline UPDRS Quartile**

QUARTILE/OTHER	ESTIMATE	STD. ERR.	P-VALUE
Base Q1 di ff	-0.8256	1.3553	0.5427
Base Q2 di ff	-0.2626	1.4042	0.8517
Base Q3 di ff	-2.1113	1.4781	0.1539
Base Q4 di ff	1.9966	1.3640	0.1439
Test for linear trend	-0.6998	1.6008	0.6622
Q4 vs. avg of Q1 to Q3	3.0631	1.6122	0.0580

In TEMPO in the Active Phase Efficacy Dataset the mean baseline UPDRS Total score was 24.5 and the quartile subgroups were  $\leq 17$ , 17- 23, 23-30.5, and  $> 30.5$ . The range was 5.5 to 75. Exploratory analysis comparing the 2mg early group to 2mg delayed group at week 52 by baseline UPDRS quartile did not reveal a linear increase in treatment group difference by baseline score,  $p=0.596$  (Table 50). In the third quartile the difference numerically favored the delayed 2mg group by 3 UPDRS points.

**Table 50 TEMPO: Week 52 2MG Early vs. 2MG Delayed Group Differences by Baseline UPDRS Quartile**

QUARTILE/OTHER	ESTIMATE	STD. ERR.	P-VALUE	LOWER 95%	UPPER 95%
Q1 di ff	-1.3844	2.1689	0.5242	-5.6672	2.8983
Q2 di ff	-4.2414	2.0501	0.0400	-8.2878	-0.1951
Q3 di ff	3.0687	2.4286	0.2079	-1.7217	7.8591
Q4 di ff	-7.0672	2.5153	0.0054	-12.0264	-2.1080
Linear trend	1.3622	2.5634	0.5958	-3.6975	6.4219
Q4 di ff vs. others	-6.2148	2.8333	0.0294	-11.8018	-0.6278

Table 51 shows the corresponding table comparing 1MG Early group to the 2mg Delayed group at Week 52. Again as for 2MG early compared to 2MG delayed, there was no significant linear trend and, in fact, 1MG early was numerically worse in the two highest baseline quartiles.

**Table 51 TEMPO: Week 52 1MG Early vs. 2MG Delayed Group Differences by Baseline UPDRS Quartile**

Label	Estimate	StdErr	P-value	Lower 95%	Upper 95%
Q1 di ff	-2.4751	2.1319	0.2472	-6.6815	1.7313
Q2 di ff	-1.1248	2.2351	0.6154	-5.5331	3.2835
Q3 di ff	1.4189	2.8086	0.6140	-4.1189	6.9567
Q4 di ff	0.9199	2.7134	0.7349	-4.4277	6.2675
Linear Trend	-2.8798	2.6266	0.2743	-8.0602	2.3006
Q4 di ff vs. others	1.6469	3.0530	0.5901	-4.3707	7.6645

In summary, overall there is no compelling, consistent evidence that the treatment effect increases with baseline UPDRS total score.

## 5. SUMMARY AND CONCLUSIONS

### 5.1 Statistical Issues and Collective Evidence

There is no replication of dose specific results between the two trials. The TEMPO study had a 2mg early group and a 1 mg early group but only a 2mg delayed group. If one nevertheless compares 2 mg delayed to 1 mg early at week 52, which it seems the sponsor had every attention of doing when they drew up the analysis for the active phase, the result is clearly not nominally significant. If one accepts the analysis plan specified LOCF analysis of Change in UPDRS from baseline at Week 52 then 2mg early group appears nominally positive compared to 2 mg delayed but it's not clear that this can be considered to support the 1 mg results in ADAGIO when the 1mg early group in TEMPO did not differentiate from the delayed group. Not to mention the facts that the analysis of the active phase was originally designated as primarily for safety and exploratory for efficacy, there wasn't a clear single primary endpoint or primary analysis population for the active phase in TEMPO, and even the TEMPO 2 mg early group is not nominally significant compared to 2mg delayed in a standard MMRM analysis, which may be more appropriate than LOCF for reasons of potential bias caused by carrying forward data within a delayed start design, as well as the other standard reasons provided recently in the statistical literature.

### 5.2 Conclusions and Recommendations

If we consider that one of the two studies was primarily designed to assess symptomatic benefit and was analyzed for evidence of delaying disease progression, essentially as an afterthought, then at best we would be in the situation of one study plus confirmatory evidence. However, neither study is robustly positive (noted inconsistencies across subgroups) and that doesn't even consider the complicating issue of the odd dose response pattern. Therefore, all things considered, there doesn't seem to be consistent compelling evidence of a delaying of Parkinson's disease progression provided by these two controlled Rasagiline trials.

## I. APPENDIX 1: Post-Hoc Natural History Staggered Start Analysis

### a. Background and Sponsor's Results

The following assumptions are made when fitting the statistical model described below, using the NHSS method to estimate NHE, and interpreting the NHE in the context of disease modification:

1. A linear model adequately fits the responses (changes from baseline) over time. If there are more than 3 measurements that are made after the symptomatic effect has been fully established, then a quadratic term should be considered (this leads to a slight modification of the NHE given below). If there are 3 measurements made in this time frame, then a linear model is recommended since a quadratic model is not stable. If 2 or fewer measurements are made in this timeframe, then this model is not appropriate to fit.
2. All treatment effects that happen quickly are symptomatic effects and all treatment effects that accumulate over a longer time are disease modifying effects. All symptomatic effects have been fully established prior to the first time point for which data are included in the model. If there are data collected prior to the time when symptomatic effects may have not been fully established, they will be excluded from the analysis.
3. There is a linear relationship between baseline clinical score and the symptomatic effect.

Furthermore, the slope describing this cross-sectional relationship is the same as that describing the unidentifiable longitudinal relationship between evolving clinical score and the symptomatic effect over time. Strictly speaking, it also assumes that the clinical score is an accurate representation of the true disease state of the patient, without measurement error. It is also assumed that the symptomatic effect does not depend on other time-related factors such as age, chronological time, and increasing plasma concentration of drug over time.

Let  $x$  be an indicator of treatment assignment ( $x = 1$  if the patient is assigned to treatment,  $x = 0$  if assigned to placebo) and let  $y_0$  be the centered baseline value of the clinical outcome (difference between a patient's baseline value and the overall baseline mean). The model for the mean change from baseline in the clinical outcome at post-randomization time  $k$ , denoted  $t_k$  and measured in years for simplicity, can be written as a simple linear model:

$$\Delta y_{ik} = \alpha_0 + \alpha_1 x_i + \tau_0 y_{i0} + \tau_1 x_i y_{i0} + \beta_0 t_k + \beta_1 x_i t_k + \gamma_0 y_{i0} t_k + \gamma_1 x_i y_{i0} t_k + e_{ik}$$

where  $i$  indicates the patient number.

In the model above,  $\alpha_1$  and  $\tau_1$  correspond to symptomatic effects, and  $\beta_1$  and  $\gamma_1$  correspond to disease modifying effects but also include effects due to the changing magnitude of symptomatic effects over time. The parameter  $\beta_0$  is the slope of the placebo group and  $\beta_1$  is the difference in the slopes of the treatment and placebo groups for a patient of average severity at baseline. The parameter  $\gamma_0$  is the change in the slope of the placebo group associated with a one point increase in severity at baseline, and  $\gamma_1$  is the change in the difference in slopes between the treatment and placebo groups associated with a one point increase in severity at baseline. Let  $\varphi$  be the difference in the slopes between the treatment group and the placebo group due solely to a

disease modifying effect, for a patient of average severity at baseline, reported as points per year. The slope of the treatment group is  $\beta_0 + \beta_1$  and the slope of the treatment group excluding symptomatic effects, i.e., the mean change in the “true” disease severity per year, is  $\beta_0 + \varphi$ .

We want to estimate  $\varphi$  by taking the difference in slopes between the treatment group and the placebo group ( $\beta_1$ ) and subtracting the impact on the slope of the change in symptomatic effect over time. The NHE estimates the disease modification effect,  $\varphi$ , for an “average” patient in the study, so the parameter estimates of  $\gamma_0$  and  $\gamma_1$  are not incorporated in the NHE. The symptomatic effect over time is estimated by using the slope of the linear relationship between baseline score and the short-term symptomatic effect. This slope,  $\tau_1$ , is the average change in the magnitude of the symptomatic effect of treatment per one unit increase in baseline severity score. According to Assumption #3 above, this is also equal to the mean change in symptomatic effect that would occur in an individual patient due to a one unit increase in the “true” disease severity over time. Since the mean change in “true” disease severity per year in the treatment group is  $\beta_0 + \varphi$ , the mean change per year in symptomatic effects due to a change of  $\beta_0 + \varphi$  units of “true” disease severity is equal to  $(\beta_0 + \varphi) \times \tau_1$ . So  $(\beta_0 + \text{NHE}) \times \tau_1$  is an estimate of the amount that changing symptomatic effects are contributing to the slope difference between the active and placebo groups over time.

The NHE is calculated by taking the difference in slopes between the treatment and placebo groups and then subtracting the estimated symptomatic contribution to this slope difference:

$$\text{NHE} = \hat{\beta}_1 - (\hat{\beta}_0 + \text{NHE}) \times \hat{\tau}_1.$$

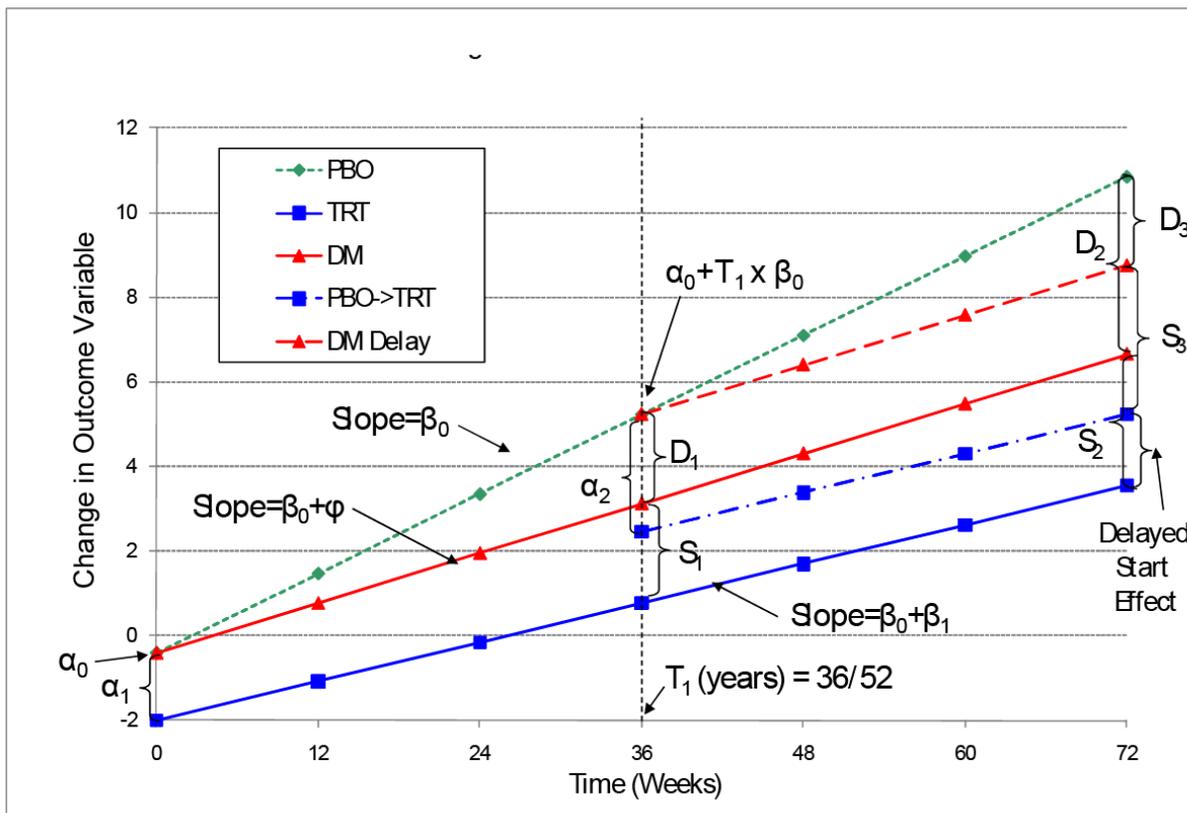
One then solves for NHE:

$$\text{NHE} = \frac{\hat{\beta}_1 - \hat{\beta}_0 \hat{\tau}_1}{1 + \hat{\tau}_1}$$

Since time is measured in years (for simplicity), the NHE can be interpreted as the number of points per year of treatment-related benefit (relative to placebo) due to disease modification, or specifically based on the proposed definition of disease modification, the treatment benefit that is accumulating over time. The standard error of this estimate can be estimated using a bootstrap procedure: resampling with replacement from the original data set and calculating the standard error (SE) of the distribution of the NHE from this resampling procedure. If  $\tau_1$  is zero, then the symptomatic effect does not depend on disease severity, and the NHE estimate is just the difference in slopes over 1 year. If  $\tau_1$  is negative, then as the disease gets worse, the symptomatic effect gets larger and the slope difference must be decreased in order to estimate the disease modifying effect. Ninety-five percent confidence intervals for the NHE are then obtained as  $\text{NHE} \pm 1.96 \times \text{SE}$  after verifying that the bootstrap distribution of the NHE is approximately normally distributed.

The slope of the DM effect is  $\beta_0 + \varphi$ , since  $\varphi$  represents a change in the slope (decline rate) from the placebo group (see Figure 1). The observed disease modification effect size can be represented as a percent reduction in decline, calculated as  $\text{NHE} / \beta_0$ .

S1 is the symptomatic effect for the Early Treatment Group at the end of the first phase  
D1 is the disease modifying effect for the Early Treatment Group at the end of the first phase  
S2 is the symptomatic effect for the Early Treatment Group at the end of the second phase  
D2 is the disease modifying effect for the Early Treatment Group at the end of the second phase  
S3 is the symptomatic effect for the Delayed Start Group at the end of the second phase  
D3 is the disease modifying effect for the Delayed Start Group at the end of the second phase.  
The Delayed Start effect is estimated as  $(D2+S2) - (D3+S3)$ , and is intended to be an estimate of D1. If the total incremental effect obtained during the second phase (the entire symptomatic effect and the disease modifying effect acquired during the second phase) is the same for the early start group  $[(D2+S2) - D1]$  and the delayed start group  $(D3+S3)$ , then the expected value of the Delayed Start estimate is equal to D1 and the Delayed Start estimate is an unbiased estimate of the disease modifying effect at the end of the first phase. If this is not the case, then the Delayed Start estimate is biased.  
 $\tau_1$  is the difference in slopes that is due to symptomatic effects – it does not change over time or with differing baseline severity.  $\alpha_2$  is the initial symptomatic effect for the delayed start group and is bigger than  $\alpha_1$  which is the initial symptomatic effect for the early start group. Although this figure does not illustrate the effect of  $\gamma_1$ , the slope for delayed start patients may be worse or better than the slope of early start patients when  $\gamma_1$  is not equal to 0.



Note: This figure copied from the sponsor's ISE Appendix 1 page 13.

## **ANALYSES OF THE ADAGIO AND TEMPO USING THE NATURAL HISTORY STAGGERED START METHOD**

The Natural History Estimator is designed to be based on data collected from a single phase placebo-controlled study. The first phase of the ADAGIO study (0 to 36 weeks) and the first phase of the TEMPO study (0 to 26 weeks) are placebo controlled, and are therefore appropriate for this analysis. However, the ADAGIO and TEMPO studies also include a delayed start phase, and a modification to the approach would allow inclusion of data from this second phase of the study as well. Second phase data for early start patients can be included as observed, but second phase data for the delayed start patients need to be “moved back” to the beginning of the study, so that the time variable that is included in the model actually represents time on the current treatment. Models 2, 3, 6, and 7 below include these “moved back” data for the delayed start patients. In these analyses, paired and unpaired data are combined (with two observations per patient included in the first phase for delayed start patients, one for the placebo condition and one for the treatment condition). Since there is no group that received placebo during the second phase of the study, any phase effect is confounded with treatment assignment in the first phase. In addition, the patients (observations) that are included in the treated group represent patients with somewhat more progressed disease than those in the placebo group since they include patient data from the first phase in the early start treatment group as well as data from patients from the second phase in the delayed start treatment group who have now progressed for the duration of the first phase of the study. By contrast, the placebo group includes only data from the first phase in the initial randomized placebo group. Adjustment for severity score at the start of treatment should reduce or even eliminate the resulting bias. Results from these analyses will be considered in conjunction with an analysis that includes only data from the first phase of the study in order to explore the sensitivity of the models. Due to these complexities, Model 1 and 5 were considered the primary analysis because they include as much data as possible for patients who remained on the original randomized groups.

Three different analysis data sets were used: the ADAGIO study, the TEMPO study and the Integrated Analysis data set which included data from both ADAGIO and TEMPO. For all analyses, the efficacy ITT population was used which included all subjects who had post-baseline efficacy data. A linear model was fit to the four different subsamples of data for each dose group versus placebo described below in order to estimate disease modification based on data from both phases of these studies. All available data were included except data that was collected at unscheduled visits. Models were run separately for the 1 mg group versus placebo and the 2 mg group versus placebo. An unstructured covariance structure (UN) was specified for each model run on the Integrated Analysis data set as well as the ADAGIO and TEMPO data sets. It was planned that if the model was unable to converge using UN, then an ARH(1) structure or an AR(1) structure would be specified. The data sets for each model for the two different dose groups are described below and are shown in Table 52.

**Table 52 ADAGIO: Natural History Staggered Start Disease Modification Estimates**

ADAGIO	Group	Model	$\beta_0$	$\beta_1$	$\tau_1$	% Reduction in Decline from Placebo rate	NHE (points per year)	95% Confidence Interval for NHE	
<b>PBO vs. Early</b>	<b>1 mg</b>	<b>1</b>	7.39	-3.38	-0.06	<b>42.3%</b>	<b>-3.13*</b>	<b>-5.06</b>	<b>-1.19</b>
PBO vs. Pooled	1 mg	2	7.68	-3.41	-0.04	42.3%	-3.25*	-5.35	-1.15
PBO vs. All	1 mg	3	7.40	-3.50	-0.12	39.8%	-2.95*	-4.76	-1.13
First Phase	1 mg	4	7.42	-2.25	-0.10	22.4%	-1.66	-3.85	0.53
<b>PBO vs. Early</b>	<b>2 mg</b>	<b>5</b>	7.48	-4.27	-0.01	<b>56.8%</b>	<b>-4.25*</b>	<b>-5.72</b>	<b>-2.77</b>
PBO vs. Pooled	2 mg	6	7.57	-4.24	-0.07	52.9%	-4.00*	-6.02	-1.97
PBO vs. All	2 mg	7	7.48	-4.06	-0.15	46.0%	-3.44*	-5.09	-1.79
First Phase	2 mg	8	7.47	-4.13	-0.01	55.1%	-4.11*	-6.24	-1.98
PBO vs. Early Using Separate Placebo Groups	1mg	9	8.05	-4.00	-0.08	45.5%	-3.66*	-6.58	-0.74
	2 mg	10	6.96	-3.75	0.01	54.5%	-3.79*	-5.67	-1.91
PBO vs. Early Same Symptomatic Effect for both Doses	1 mg	11a	7.41	-3.60	-0.04	46.6%	-3.45*	-5.25	-1.65
	2 mg	11b	7.41	-3.88	-0.04	50.5%	-3.74*	-5.49	-1.98

\*Significant at two-sided 5% significance level.

Note: This table was copied from sponsor’s ISE appendix 1 page 19

**1 mg Group vs. Placebo:**

Model 1: “Placebo vs. Early Treatment” includes data from the first phase of the study plus continuing data from the second phase for patients who started treatment early (excludes second phase data for delayed start patients).

Model 2: “Placebo vs. Pooled” includes data from the first phase of the study plus data from the second phase for patients who were in the delayed start group. Data from the second phase for the delayed start patients was “pooled” with the data from the first phase by “moving the data back” to time 0, so that time represented the time on the current treatment. Also, the baseline values were set to the value at the time that treatment was started. This model excludes second phase data for early start patients.

Model 3: “Placebo vs. All” includes all data from Models 1 and 2.

Model 4: “First Phase Only” includes only data from the placebo controlled phase of the study.

**2 mg Group vs. Placebo:**

Model 5: Same as Model 1 above.

Model 6: Same as Model 2 above.

Model 7: Same as Model 3 above.

Model 8: Same as Model 4 above.

**Table 53 TEMPO: Natural History Staggered Start Disease Modification Estimates**

TEMPO	Group	Model	$\beta_0$	$\beta_1$	$\tau_1$	% Reduction in Decline from Placebo rate	NHE (points per year)	95% Confidence Interval for NHE	
PBO vs. Early (same as PBO vs. All)	1 mg	1 & 3	10.91	-6.05	-0.03	53.9%	-5.89*	-11.16	-0.61
PBO vs. Pooled (same as First Phase Only)	1 mg	2 & 4	10.9	-4.45	-0.04	38.4%	-4.19	-10.09	1.71
PBO vs. Early	2 mg	5	10.59	-5.75	0.09	58.2%	-6.16*	-11.83	-0.49
PBO vs. Pooled	2 mg	6	10.33	-3.28	-0.14	20.7%	-2.13	-9.40	5.13
PBO vs. All	2 mg	7	10.14	-5.42	-0.13	46.4%	-4.70	-10.31	0.90
First Phase Only	2 mg	8	10.60	-4.20	-0.06	36.0%	-3.81	-11.21	3.59
PBO vs. Early Using Separate Placebo Groups	1mg	9	Not Applicable (no Delayed Start 1 mg group)						
	2 mg	10	Same as Model 5						
PBO vs. Early Same Symptomatic Effect for both Doses	1 mg	11a	10.75	-5.70	0.02	53.8%	-5.78*	-10.35	-1.22
	2 mg	11b	10.75	-5.61	0.02	53.0%	-5.70*	-10.35	-1.04

\*Significant at two-sided 5% significance level.

Note: This table was copied from sponsor’s ISE appendix 1 page 21

**Table 54 ADAGIO + TEMPO: Natural History Staggered Start Disease Modification Estimates**

ADAGIO + TEMPO	Group	Model	$\beta_0$	$\beta_1$	$\tau_1$	% Reduction in Decline from Placebo rate	NHE (points per year)	95% Confidence Interval for NHE	
PBO vs. Early	1 mg	1	7.85	-3.44	-0.04	41.3%	-3.24*	-4.88	-1.60
PBO vs. Pooled	1 mg	2	8.03	-3.64	-0.02	44.2%	-3.55*	-5.41	-1.69
PBO vs. All	1 mg	3	7.74	-3.50	-0.10	38.9%	-3.01*	-4.77	-1.25
First Phase Only	1 mg	4	7.86	-2.58	-0.08	27.4%	-2.15	-4.40	0.09
PBO vs. Early	2 mg	5	7.89	-4.34	0.02	55.7%	-4.40*	-6.06	-2.73
PBO vs. Pooled	2 mg	6	8.01	-4.20	-0.06	49.2%	-3.94*	-5.54	-2.35
PBO vs. All	2 mg	7	7.88	-4.12	-0.13	45.0%	-3.55*	-5.30	-1.80
First Phase Only	2 mg	8	7.88	-4.22	0.004	53.8%	-4.24*	-6.12	-2.35
PBO vs. Early Using Separate Placebo Groups	1mg	9	8.33	-3.89	-0.10	40.5%	-3.38*	-5.81	-0.94
	2 mg	10	7.58	-3.00	0.04	54.3%	-4.12*	-6.01	-2.23
PBO vs. Early Same Symptomatic Effect for both Doses	1 mg	11a	7.88	-3.63	-0.02	45.3%	-3.57*	-5.21	-1.93
	2 mg	11b	7.88	-4.04	-0.02	50.5%	-3.98*	-5.58	-2.37

\*Significant at two-sided 5% significance level.

Note: This table was copied from sponsor’s ISE appendix 1 page 23

In summary, the sponsor concludes the following concerning the natural history staggered start analysis.

“The analyses of the ADAGIO trial demonstrated a consistent, statistically significant disease modifying benefit with both the 1mg and 2 mg doses of rasagiline. The results from analyses of the TEMPO trial though not always distinguishable from noise due to the smaller sample size, were consistent in direction and magnitude with those from ADAGIO. Naturally, when the data from these two trials were pooled together, a consistent disease modifying benefit was also observed.”

The sponsor only offers the following brief paragraph to explain the marked difference between the results of the primary analysis method and the natural history staggered start analysis. “Because the NHE and the Delayed Start estimates are estimating different effects, it is not surprising that they give different results. The NHE is an estimate of the disease modification or aggregating effects on drug, and the Delayed Start estimate is an estimate of the early treatment benefit which includes these aggregating effects on drug but also includes any penalty or reward for early treatment. Because we fit a linear model for these data with a constant slope over time, there is no penalty or reward for early treatment in the disease modifying effect, but only in the symptomatic effect.”

Finally, it should further be noted that these NHSS analyses are all *post-hoc* in that they were not planned prior to the unblinding of either trial and that no adjustment to the CI’s for multiple comparisons has been made.

#### **b. Reviewer’s Comments on Natural History Staggered Start Analysis**

The first problem with this natural history analysis which is irrefutable is that it is post-hoc. In addition, this analysis suffers from it being designed for a one phase study thus requiring us in the present case to discard a substantial amount of available data from the delayed start design. This also creates a group imbalance in the form of a systematic difference between the early and delayed groups in the assessment time schedule (except in the phase 1 only analysis) which may invalidate or bias the treatment group comparison. The approach also seems to suffer from the fact that the estimate it produces does not apply to the whole population but only to those with the study’s average baseline severity. The distribution of the baseline UPDRS in ADAGIO is not symmetric (the median is 19 and the mean is 20.4 in the ITT population). So 56% have a lower baseline than 20.4. For the 1mg groups in the ACTE population the median is 18.5 and the mean is 19.83 and again 55.4% have a baseline lower than the mean. It’s actually 57% below the mean for 1mg delayed and 53.8% below for 1 mg early. The patients with exactly the average baseline severity could be a small subset and it leaves one wondering about the rest of the population. For example, can we be assured that the results for the rest of the group would not be discordant with those with average baseline severity? Wouldn’t a population average be a more clinically relevant quantity?

In the typical analysis as in the hypothesis 2 analysis for a difference at week 72 the estimated overall treatment group difference can be interpreted as the average treatment difference that pertains to all study patients regardless of their particular baseline demographics and baseline disease characteristics. However, in contrast, the NHE only applies to subjects who had the average baseline UPDRS score. There may be very few patients exactly at average baseline, i.e., to whom the NHE applies. Furthermore, the sponsor’s presentation gives us no information on how much the NHE might change if we were to compute a corresponding NHE estimate for a baseline UPDRS score different from the study average.

This also implies that pooling the ADAGIO and TEMPO data for this Natural History Staggered Start analysis is particularly questionable. In a meta analysis, one typically pools studies of

patients with similar demographics but the baseline UPDRS was about 5 points higher on average for the TEMPO population than the ADAGIO population. Therefore, the average baseline severity in the pooled dataset which the natural history estimate would correspond to may not represent either component study well.

The natural history model assumes a linear slope relationship between change from baseline and time such that the slope increases linearly as a function of the baseline UPDRS score. This can be shown to imply a nonlinear relationship between change in UPDRS and time unless the dependence of the slope on the baseline severity is very weak. Suppose an individual with average baseline severity, denoted  $y_{i0}$ , has a slope of  $\beta$  for the relationship between change and time. Then in time  $1/\beta$  this individual will have the same UPDRS as another person with a baseline of  $y_{i0}+1$ . All else being the same why shouldn't the patient's (with a baseline of  $y_{i0}$ ) slope now after time  $1/\beta$  become  $\beta + \gamma(y_{i0}+1 - y_{i0}) = \beta + \gamma$  as it was for the other person? Unless  $\gamma$  is very small or  $1/\beta$  is very long this constant subject slope seems problematic. Doesn't it contradict the other person's slope being  $\beta + \gamma$  and not  $\beta$ ? We can use this same reasoning for any patients separated by 1 UPDRS unit at baseline. If we assume across all baseline UPDRS values, all else being equal, that after progressing one unit a patient's slope becomes equal to that of a patient with a UPDRS 1 point higher at baseline then we arrive at a nonlinear relationship between UPDRS and time. This seems to contradict the NHE model's assumed constancy of slope over time for each individual. Also, in general, the longer the scheduled duration of a patient's participation in the trial the less likely linearity is to hold throughout that duration.

If slope and intercept of the UPDRS relationship with time depend on baseline UPDRS score then perhaps the duration of the symptomatic effect also would depend on the baseline UPDRS score. This reviewer expects if we don't get duration of symptomatic effect right, i.e., don't exclude from the analysis all data before the symptomatic effect is fully established then the type I error will likely be inflated. In particular, we could improperly conclude a symptomatic effect is disease modifying if we are not careful with the natural history analysis.

The differential treatment effect by Gender as suggested by the primary analysis of ADAGIO applied to Gender subgroups also seems to exist for the Natural History Staggered Start Analysis (Table 55).

**Table 55 ADAGIO: 1MG NHE Model 1 Analysis by Gender**

GENDER	$\beta_0$	$\beta_1$	$\tau$	% REDUCTION	NHE	95% LOWER	95% UPPER	P- VALUE
Males(N=537)	7.88	-2.85	-0.15	25.14	-1.98	-4.21	0.25	0.0815
Females(N=337)	7.38	-5.26	-0.03	72.10	-5.32	-7.59	-3.05	0.0001

It is important to assess the agreement between the model and the observed data for this model since it assumes that the symptomatic effect increases linearly as a function of the baseline across the whole range of the baseline UPDRS. In fact, this cannot be the case at both limits of the UPDRS because the UPDRS can't go beyond its defined range of scores. The sponsor does not seem to have evaluated the fit of this model aside from adding a quadratic term and checking its significance which doesn't really inform us about the overall fit of the linear model. One

other implicit assumption of the model is that the variance does not depend on the baseline UPDRS. It is not uncommon in Statistics to find that the variance increases as the mean increases. This could happen in the present case for which the mean is assumed to increase with the baseline score. Some evidence that this may in fact be the case for NHE model 1 was found when it was re-fit allowing the variances to vary by baseline UPDRS quartile. For example, in the following table of variances by week and baseline quartile Columns 3 and 4, corresponding to quartiles 3 and 4, are consistently bigger than the corresponding entries of column 1 or 2, corresponding to quartiles 1 and 2. Also, aside from week 66 column 2 is always bigger than column 1 for a given week, and column 3 (as well as 4) is always bigger than column 2 for a given week. Therefore it seems this implicit assumption of the model may be suspect.

**Table 56 ADAGIO NHE Model 1 Residual Variance by Week and Baseline Quartile**

<b><u>Week</u></b>	<b>Baseline Quartile</b>			
	<b><u>1</u></b>	<b><u>2</u></b>	<b><u>3</u></b>	<b><u>4</u></b>
12	14.6	23.6	31.6	39.2
24	17.3	34.6	40.2	55.4
36	18.1	40.4	54.1	71.0
42	26.5	31.0	51.1	79.9
48	32.4	35.6	84.7	83.8
54	32.2	45.7	74.5	82.4
60	33.6	47.2	93.6	74.5
66	50.5	48.0	103.5	81.3
72	51.4	62.1	105.0	112.6

In summary, first, this natural history staggered start model is post-hoc. There doesn't seem to be any justifiable reason to elevate it's stature. The associated model involves many assumptions some of which are not even testable. In order to apply this model designed for a single phase trial to the delayed start design trial we need to discard an entire phase of data for at least one treatment group. Finally, the sponsor didn't provide any compelling explanation for the big difference in conclusions between the primary analysis method and the analysis based on this natural history staggered start model. This needs to be investigated because if it was due to an invalid assumption then the reported results of the natural history model may be biased. For these reasons, the natural history staggered start analysis results are considered inconclusive and exploratory by this reviewer.

## **II. Appendix 2: References**

1. Parkinson Study Group. A controlled trial of rasagiline in early Parkinson disease: the TEMPO Study. *Arch Neurol.* 2002 Dec;59(12):1937-43.
2. Parkinson Study Group. A Randomized Placebo-Controlled Trial of Rasagiline in Levodopa-Treated Patients With Parkinson Disease and Motor Fluctuations: The PRESTO Study. *Arch Neurol.* 2005;62:241-248.
3. Rascol, D J Brooks, E Melamed, W Oertel, W Poewe, F Stocchi, E Tolosa, for the LARGO study group. Rasagiline as an adjunct to levodopa in patients with Parkinson's disease and motor fluctuations (LARGO, Lasting effect in Adjunct therapy with Rasagiline Given Once daily, study): a randomised, double-blind, parallel-group trial. *Lancet* 2005; 365: 947–54.





U.S. Department of Health and Human Services  
Food and Drug Administration  
Center for Drug Evaluation and Research  
Office of Biostatistics

Statistical Review and Evaluation  
CLINICAL STUDIES

**NDA/Serial Number:** NDA021641  
**Drug Name:** Rasagiline Mesylate  
**Indication(s):** PD Disease Modifying  
**Applicant:** Teva Pharmaceutical Industries Ltd.  
**Date of Document:** Dec 23, 2010  
**Biometrics Division:** Division 1 (HFD-710)  
**Statistical Reviewer:** Ohidul Siddiqui, Ph.D  
**Concurring Reviewers:** Kun Jin, Ph.D; Jim Hung, Ph.D  
**Medical Division:** HFD-120  
**Medical Reviewer:** Leonard P Kapcala, MD  
**Project Manager:** Stacy Metz  
**Keywords:** *Clinical studies, NDA review, statistical review, PD disease modifying benefit, disease modeling*

## **Table of Contents**

1. EXECUTIVE SUMMARY .....	3
1.1. CONCLUSIONS AND RECOMMENDATIONS .....	3
1.2. BRIEF OVERVIEW OF REVIEWED CLINICAL STUDIES.....	3
1.3. STATISTICAL ISSUES AND FINDINGS.....	4
2. INTRODUCTION .....	4
2.1 OVERVIEW .....	4
2.2 DATA SOURCES .....	10
3. STATISTICAL EVALUATION .....	10
3.1 DEMOGRAPHIC AND BASELINE CHARACTERISTICS.....	10
3.2 EFFICACY EVALUATION .....	11
4. SUBGROUP ANALYSIS .....	22
5. SUMMARY AND CONCLUSIONS .....	24

## 1. EXECUTIVE SUMMARY

### 1.1. CONCLUSIONS AND RECOMMENDATIONS

The Sponsor claims that the 1 mg dose (in the ADAGIO study) and the 2 mg dose (in the TEMPO study) of rasagiline could demonstrate a disease modifying benefit in patients with early untreated idiopathic Parkinson's disease. This reviewer's analyses do not support the claim for a disease modifying benefit associated with either dose of rasagiline based on the primary protocol specified analyses or when sensitivity/secondary analyses are applied to the study data sets.

### 1.2. BRIEF OVERVIEW OF REVIEWED CLINICAL STUDIES

The Sponsor has submitted the efficacy findings of two studies (ADAGIO study and TEMPO study) to include a claim of the slowing of clinical progression of Parkinson's disease to the currently labeled indication for Azilect.

ADAGIO was a multicenter, double-blind, randomized-start, placebo-controlled (PC), parallel-group Phase IIIb study to assess rasagiline as a disease modifying therapy in early untreated idiopathic Parkinson's disease (PD) subjects. The study consisted of 2 phases: phase I - a 36-week double-blind, PC phase, and phase II - a 36-week double-blind, active-treatment (AC) phase. Subjects were randomized in a 1:1:1:1 ratio into one of the following four treatment groups: (i) 1 mg/day rasagiline during Phase I and Phase II (1 mg early start), (ii) 2 mg/day rasagiline during Phase I and Phase II (2 mg early start), (iii) Placebo during Phase I followed by 1 mg/day rasagiline during Phase II (1 mg delayed start), and (iv) Placebo during Phase I followed by 2 mg/day rasagiline during Phase II (2 mg delayed start).

TEMPO was a multicenter, double-blind, placebo-controlled, parallel group, phase III clinical study to assess the efficacy, tolerability and safety of two doses of rasagiline mesylate in early stage PD patients not treated with levodopa. The TEMPO study also consisted of 2 phases: Phase I - a 26-week double-blind, placebo-controlled (PC) phase, and Phase II - a 26-week double-blind, active-treatment (AC) phase. Subjects were randomized in a 1:1:1 ratio into one of the following three treatment groups: (i) 1 mg/day rasagiline during Phase I and Phase II (1 mg early start), (ii) 2 mg/day rasagiline during Phase I and Phase II (2 mg early start), and (iii) placebo during Phase I followed by 2 mg/day rasagiline during Phase II (2 mg delayed start). There was no delayed-start group for the 1 mg dose.

#### Statistical Analysis

For testing Hypothesis #1, a Linear Mixed Model with random intercept and slope was used. The changes from baseline in total UPDRS score for each post baseline visit in the PC phase were used as dependent measures. For testing Hypothesis #2 (i.e., Superiority of Early over Delayed Start of rasagiline at the end of last week in AC phase, a Repeated Measures model was used. For testing Hypothesis #3 (Slopes Non-Inferiority of Early Start over Delayed Start

in the AC Phase), a Linear Mixed Model with random intercept and slope was used. In the analyses for testing Hypothesis#2 and Hypothesis#3, the changes from baseline total UPDRS scores from each post baseline visit in the AC phase were used as dependent measures.

The results of the TEMPO study were re-analyzed in the same manner as the ADAGIO study. Since there was no 1 mg delayed-start group in the study, the 1 mg early-start group was only compared with Placebo at Phase I.

Three hierarchical statistical hypothesis tests were applied the primary efficacy endpoint. The Hochberg's Step-Up Bonferroni method for multiple comparisons between treatment groups, in combination with the hierarchical method for the three hypotheses testing, are used to maintain the experiment-wise type I error of 5%. If the first null hypothesis is not rejected for one of the doses at an alpha level of 5%, then the other dose is tested using an alpha level of  $5\%/2 = 2.5\%$ . Each statistically significant dose, as determined by the test of hypothesis #1, is further tested for hypothesis #2. Each statistically significant dose, as determined by the test of hypothesis #2, is further tested for hypothesis #3.

### 1.3. STATISTICAL ISSUES AND FINDINGS

It is statistically challenging to try to determine if there is a disease modifying benefit associated with any drug and it is more of a challenge to differentiate symptomatic benefits from a disease modifying benefit. In the ADAGIO study, it was assumed that the full rasagiline symptomatic effect is present by week 12, and that there is a linear relationship for the data points from week 12 data onwards. However, the analysis of ADAGIO data did not support these assumptions. The data indicates there is a nonlinear relationship for data between weeks 12 and 36 hence; the symptomatic benefits were not fully established at the end of week 12. The significances of the slope differences in the original analysis were mainly due to the nonlinear relation between the data from weeks 12 to 36. Since different drugs might require different amounts of time on drug to fully establish their symptomatic benefits, it is statistically challenging to separate out symptomatic benefits from the disease modifying benefits for drugs in development.

Since trials designed to study a disease modifying potential in a chronic illness require longer observation periods compared to typical efficacy trials, dealing with missing data due to patient dropout in the PC and AC phases also presents a statistical challenge when evaluating the disease modifying potential of a drug.

## 2. INTRODUCTION

### 2.1 OVERVIEW

The Sponsor has submitted efficacy findings of two studies to support a claim for the slowing of clinical progression of Parkinson's disease in the current labeled indication for Azilect. The

proposed change is supported by the data presented from two clinical trials: TVP-1012/500 (ADAGIO) - “A Multicenter, Double-Blind, Randomized Start, Placebo-Controlled, Parallel-Group Study to Assess rasagiline as a Disease Modifying Therapy in Early Parkinson’s Disease Subjects” and TVP-1012/232 (TEMPO) - “A multicenter, double-blind, placebo controlled, parallel group, phase III clinical study for the efficacy, tolerability and safety of two doses of rasagiline mesylate in early Parkinson’s Disease (PD) patients not treated with levodopa”. Figures 1 and 2 list the design features of the studies.

### ADAGIO Study Design

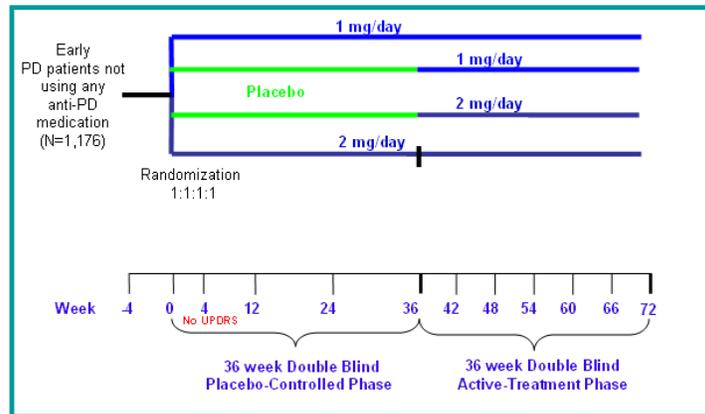


Figure 1. ADAGIO Study Design  
 Source: Summary of clinical efficacy report

ADAGIO was a multicenter, double-blind, randomized-start, placebo-controlled (PC), parallel-group Phase IIIb study to assess rasagiline as a disease modifying therapy in early untreated idiopathic Parkinson’s disease subjects. The study consisted of 2 phases: phase I - a 36-week double-blind, PC phase, and phase II - a 36-week double-blind, active-treatment phase. Subjects were randomized in a 1:1:1:1 ratio into one of the following four treatment groups: (i) 1 mg/day rasagiline during Phase I and Phase II (1 mg early start), (ii) 2 mg/day rasagiline during Phase I and Phase II (2 mg early start), (iii) Placebo during Phase I followed by 1 mg/day rasagiline during Phase II (1 mg delayed start), and (iv) Placebo during Phase I followed by 2 mg/day rasagiline during Phase II (2 mg delayed start).

## TEMPO Study Design

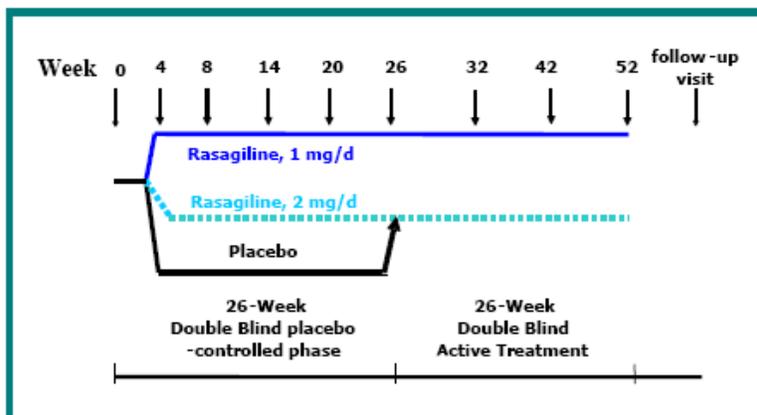


Figure 2. TEMPO Study Design

Source: Summary of clinical efficacy report

TEMPO was a multicenter, double-blind, placebo-controlled, parallel group, phase III clinical study to assess the efficacy, tolerability and safety of two doses of rasagiline mesylate in early stage PD patients not treated with levodopa. The TEMPO study consisted of 2 phases: Phase I - a 26-week double-blind, placebo-controlled phase, and Phase II - a 26-week double-blind, active-treatment phase. Subjects were randomized in a 1:1:1 ratio into one of the following three treatment groups: (i) 1 mg/day rasagiline during Phase I and Phase II (1 mg early start), (ii) 2 mg/day rasagiline during Phase I and Phase II (2 mg early start), and (iii) placebo during Phase I followed by 2 mg/day rasagiline during Phase II (2 mg delayed start). There was no delayed-start group for the 1 mg dose.

### ADAGIO Study

The primary objective of the ADAGIO study was to assess whether rasagiline has a disease modifying effect in patients with early Parkinson's disease. Three hierarchical hypotheses were tested based on changes from baseline in the total Unified Parkinson's Disease Rating Scale (UPDRS) score (Parts I, II and III)- (i) Hypothesis #1: Superiority of rasagiline over placebo in the slopes of UPDRS in Phase I ( weeks 12-36 weeks), (ii) Hypothesis #2: Superiority of Early over Delayed Start at Week 72 (using repeated measures based on Weeks 48-72), and (iii) Hypothesis #3: Slopes Non-Inferiority of Early Start over Delayed Start in Phase II of the studies (Weeks 48-72).

### Statistical Analysis

For testing Hypothesis #1, a Linear Mixed Model with random intercept and slope was used. The change from baseline in the total UPDRS score at each post-baseline visit in the PC Phase was used as a dependent measure. The model included treatment group, continuous week in trial by treatment interaction, center and baseline Total UPDRS score as fixed effects, and individual subject intercept and the week effects as random effects. An "unstructured" (UN) covariance matrix between the intercept and slopes estimates was used. Two comparisons were derived from the model: (i) slope difference of rasagiline 1 mg group from the placebo

group and (ii) slope difference of rasagiline 2 mg group from the placebo group. In this analysis, all available post-baseline observations in the PC Phase of the trial were analyzed (ITT efficacy data analysis set). The placebo groups for rasagiline 1 mg and 2 mg were combined to one placebo group.

For testing Hypothesis #2 (Superiority of Early over Delayed Start of rasagiline at the end of last week, a Repeated Measures Model was used. The change from baseline in total UPDRS at each post-baseline visit in the AC Phase was used as a dependent measure. The model included categorical week in trial by treatment interaction, center, and baseline Total UPDRS score as fixed effects. An “unstructured” (UN) covariance matrix for repeated observations within subjects was used. The least square mean (LSMEAN) at week 72 of the change from baseline in Total UPDRS were compared (i) between the rasagiline 1 mg early-start group and the rasagiline 1 mg delayed-start group, and (ii) between the rasagiline 2 mg early-start group and the rasagiline 2 mg delayed-start group. In this analysis, observations of all subjects entering the active phase with at least 24 weeks of treatment during the PC Phase and at least one available Total UPDRS measurement during the active treatment phase from weeks 48, 54, 60, 66 or 72, were analyzed (ACTE data analysis set).

Hypothesis #3 (Slopes Non-Inferiority of Early Start over Delayed Start in the Active Phase), is stated as follows:

H0:  $\text{Slope}_{(\text{Early Start Group})} - \text{Slope}_{(\text{Delayed Start Group})} > 0.15$

HA:  $\text{Slope}_{(\text{Early Start Group})} - \text{Slope}_{(\text{Delayed Start Group})} \leq 0.15$

Slope is the model estimate of the change from baseline in total UPDRS per week. In this analysis, observations of all subjects entering the active phase with at least 24 weeks of treatment during the PC Phase and at least one available Total UPDRS measurement during the active treatment phase from weeks 48, 54, 60, 66 or 72 were analyzed (ACTE data analysis set). The statistical model was a Linear Mixed Model with random intercept and slope. The model included treatment group, continuous week in trial by treatment interaction, center and baseline Total UPDRS score as fixed effects, and the individual subject intercept and the week effects as random effects. An “unstructured” (UN) covariance matrix between the intercept and slopes estimates was used. A noninferiority test for the difference in slopes between the treatment groups was performed. The one sided 95% Confidence Intervals (CI) were calculated for the difference between the slopes of the rasagiline 1 mg early-start group and the rasagiline 1 mg delayed-start group and between the slopes of the rasagiline 2 mg early-start group and the rasagiline 2 mg delayed-start rasagiline group. The inferiority null hypothesis of the early-start group slope over the delayed-start group slope was to be rejected if the upper limit of the one sided 95% CI for the difference in slopes did not cross the noninferiority margin of 0.15 UPDRS points per week.

### TEMPO study

To make a similar comparison for the results of both trials, the efficacy data from the TEMPO study was re-analyzed in the same manner as the ADAGIO study. Since there was no 1 mg delayed-start group, the 1 mg early-start group was only compared with Placebo in Phase I.

## Data Analysis Sets

Following data sets were defined for the efficacy analyses in the ADAGIO study and were adapted to the reanalysis of the TEMPO study:

Intent-to-Treat Data Analysis Set (ITT) consists of all subjects randomized with at least one post-baseline measurement. In accordance with the ITT principle, subjects are kept in their originally assigned treatment group.

In the primary efficacy analysis for testing hypothesis #1, Intent-to-Treat Data Analysis Sets (ITT) were used. ITT Data Analysis Sets consisted of subjects who had at least one of the UPDRS assessments performed at weeks 12, 24 and 36 as mandated in the ADAGIO protocol, and at least one of the UPDRS assessments performed at weeks 14, 20 and 26 in TEMPO.

For the analyses of hypotheses #2 and #3, Active Efficacy Data Analysis Sets (ACTE) were used. Active Efficacy Data Analysis Sets (ACTE) consisted of all subjects entering the active-treatment phase of the trial with at least 24 weeks of treatment during the PC phase of ADAGIO or TEMPO, and at least one available Total UPDRS measurement during the active-treatment phase from Week 48 onwards in ADAGIO and from week 42 in TEMPO.

## Multiplicity Adjustment

Three hierarchical statistical hypothesis testing applied on the primary efficacy endpoint. The Hochberg's Step-Up Bonferroni method for multiple comparisons between treatment groups, in combination with the hierarchical method for the three hypotheses testing, are used to maintain the experiment-wise type I error of 5%. If the first null hypothesis is not rejected for one of the doses at an alpha level of 5%, then the other dose is tested using an alpha level of  $5\%/2 = 2.5\%$ . Each statistically significant dose, as determined by the test of hypothesis #1, is further tested for hypothesis #2. Each statistically significant dose, as determined by the test of hypothesis #2, is further tested for hypothesis #3.

## Disposition of Patients

The majority of patients in both studies completed their respective studies as planned (Table 1). Only 7.1% of the 1174 subjects who were randomized to the ADAGIO study and received at least one dose of study medication terminated prematurely during the PC phase. In Phase II of ADAGIO study, only 12.6% subjects terminated from the study before the study endpoint. In both phases of ADAGIO study, the most common reason for premature termination was due to adverse events (AEs) and was similar (not statistically significant) across groups.

In the TEMPO study, only 5.4% and 5.3% subjects were terminated in Phase I and Phase II, respectively. In both phases of the TEMPO study, the most common reason for premature termination was due to adverse events (AE) with similar frequency (not statistically significant) across groups.

Table 1. Patient Disposition

ADAGIO: Subject Disposition during Placebo-Controlled Phase										
	1 mg Delayed Start		1 mg Early Start		2 mg Delayed Start		2 mg Early Start		All	
	N	%	N	%	N	%	N	%	N	%
Received Study Medication	298	100	288	100	295	100	293	100	1174	100
Entered into Active Phase After Completing PC Phase	211	70.8	245	85.1	216	73.2	242	82.6	914	77.9
Early Transfer from Placebo Phase to Active Phase	59	19.8	28	9.7	59	20.0	31	10.6		15.1
Premature Termination during PC Phase	28	9.4	15	5.2	20	6.8	20	6.8	83	7.1
ADAGIO: Subject Disposition during Active-Treatment Phase										
	1 mg Delayed Start		1 mg Early Start		2 mg Delayed Start		2 mg Early Start		All	
	N	%	N	%	N	%	N	%	N	%
Entered into Active Phase	270	100	273	100	275	100	273	100	1091	100
Completed the Study	231	85.6	238	87.2	241	87.6	244	89.4	854	87.4
Premature Termination during Active Phase	39	14.4	35	12.8	34	12.4	29	10.6	137	12.6
TEMPO: Subjects Disposition during Placebo-Controlled Phase										
	1 mg		2 mg		Placebo		All			
	N	%	N	%	N	%	N	%	N	%
Received study Medication	134	100	132	100	138	100	404	100		
Completed PC Phase without need for additional PD Therapy	111	82.8	105	79.5	112	81.2	328	81.2		
Early Transfer to Active Phase due to need for additional PD Therapy	14	10.4	19	14.4	21	15.2	54	13.4		
Premature Termination During Placebo-Controlled Phase	9	6.7	8	6.1	5	3.6	22	5.4		
TEMPO: Subject Disposition during Active-Treatment Phase										
	1 mg Early Start		2 mg Early Start		2 mg Delayed Start		All			
	N	%	N	%	N	%	N	%	N	%
Entered into Active Phase	124	100.0	124	100.0	132	100	380	100		
Completed the Study	120	96.8	118	95.2	122	92.4	360	94.7		
Premature Termination during Active Phase	4	3.2	6	4.8	10	7.6	20	5.3		

Source: study reports

Table 2 lists the number of subjects in each data set for the statistical analysis. In ADAGIO study, only 10 subjects were excluded from the analysis data set for hypothesis #1. An additional 168 subjects were excluded from the ACTE analysis set for hypotheses #2 and #3, of whom, 57 subjects had early transfer to the active phase prior to week 24. Overall, the ACTE data analysis set included 996 subjects (84.8% out of the study ITT data set).

In TEMPO, 21 subjects who prematurely terminated from the placebo-controlled phase of the trial did not have UPDRS data from Week 14 and onwards and were excluded from the analysis data set for hypothesis #1. The analysis for hypotheses #2 and #3 was based on a data set of 279 subjects.

Table 2: Number of Subjects in Data Analysis Sets

<b>ADAGIO</b>	1 mg Delayed Start	1 mg Early Start	2 mg Delayed Start	2 mg Early Start	All
Randomized	300	288	295	293	1176
Study ITT	298	288	295	293	1174
Efficacy ITT for hypothesis #1	295 (99.0%)	286 (99.3%)	293 (99.3%)	290 (99%)	1164 a (99.2%)
ACTE for hypotheses #2 and #3	238 (79.9%)	251 (87.2%)	249 (84.4%)	258 (88%)	996 b (84.8%)
<b>TEMPO</b>					
	Early 1 mg	Early 2mg	Delayed 2mg	All	
Randomized	134	132	138	404	
Study ITT	134 (100.0%)	132 (100%)	138 (100%)	404(100%)	
Efficacy ITT for hypothesis #1	125 (93.3%)	123 (93.2%)	135 (97.8%)	383 (94.8%)	
ACTE for hypotheses #2 and #3	96 (71.6%)	89 (67.4%)	94 (68.1%)	279 (69.1%)	

Source: study reports

## 2.2 DATA SOURCES

The study reports and SAS data sets are available internally as follows:

<\\cdsesub1\evsprod\NDA021641\0030\m5\datasets>

## 3. STATISTICAL EVALUATION

### 3.1 DEMOGRAPHIC AND BASELINE CHARACTERISTICS

Table 3 lists the distribution of demographic characteristics across the two studies. Demographic characteristics were similar in both studies, except for the fact that TEMPO was conducted only in North America.

Distribution of baseline UPDRS, baseline Modified Hoehn and Yahr Scale and time since diagnosis of PD in ADAGIO and TEMPO is presented in Table 4. Baseline characteristics were comparable between treatment groups within each study. The patients in ADAGIO were at an earlier stage of their PD, as reflected by time from PD diagnosis and the baseline total UPDRS score.

Table 3. Distribution of Demographic Characteristics across Studies

Demographic Parameters	ADAGIO (n=1174)*	TEMPO (n=404)
Gender		
Female	n (%) 457 (38.9%)	147 (36.4%)

Male	n (%)	717 (61.1%)	257 (63.6%)
Age (Years)	Mean (SD)	62.2 (9.6)	60.8 (10.8)
Age (Range Years)		31-81	32-92
Race			
Caucasian	n (%)	1147 (97.7%)	383 (94.8%)
Geographical Region			
North America (USA/Canada)	n (%)	494 (42.1%)	404 (100%)
Europe/Argentina/Israel	n (%)	680 (57.9%)	-

\*Two subjects out 1176 randomized withdrew consent prior to receiving any study drug.

Source: study reports

Table 4. Distribution of Baseline PD Characteristics across Studies

		ADAGIO	TEMPO
Time from Diagnosis (days)	N	1174	404
	Mean	137.6	367.75
	SD	140.9	446.17
	Min	1	5
	Median	83.0	218.00
	Max	547	3868
Baseline Total UPDRS	N	1174	404
	Mean	20.39	25.03
	SD	8.52	10.84
	Min	3.0	5.50
	Median	19.0	23.00
	Max	53.0	75.00
Baseline Modified Hoehn and Yahr Scale	N	1174	404
	Mean	1.51	1.86
	SD	0.48	0.48
	Min	1.0	1.00
	Median	1.5	2.00
	Max	2.5	3.00

Source: study reports

### 3.2 EFFICACY EVALUATION

#### The Sponsor's Analysis Results

##### Hypothesis #1: Primary Analysis

Table 5 lists the slope comparisons in the PC phase for ADAGIO and TEMPO. In ADAGIO, the comparisons between placebo to the 1 mg and to the 2 mg early start rasagiline groups were different from zero ( $p\text{-value} \leq 0.0133$ ). The negative slope differences between rasagiline groups vs. placebo in the PC phase means that there was a slower rate of disease progression for the patients randomized to either rasagiline group as compared to patients randomized to placebo. In the TEMPO study, both comparisons between placebo, and the 1 mg or 2 mg rasagiline groups were not statistically significantly ( $p\text{-value} \geq 0.1342$ ) different from zero.

Table 5: Comparison of Changes per Week (Slope) for Rasagiline vs. Placebo During the PC Phase (ADAGIO: Weeks 12, 24, 36; TEMPO: Weeks 14, 20, 26) – Efficacy ITT Data Analysis Set

Comparison	ADAGIO					TEMPO				
	Est	SE	P-Value	Lower 95% CI Limit	Upper 95% CI Limit	Est	SE	P-Value	Lower 95% CI Limit	Upper 95% CI Limit
1 mg-Placebo Slope Difference	-0.046	0.019	0.0133	-0.083	-0.010	-0.085	0.057	0.1342	-0.197	0.026
2 mg-Placebo Slope Difference	-0.072	0.019	0.0001	-0.109	-0.036	-0.083	0.057	0.1475	-0.197	0.030

Source: study reports

Est: Estimate

Linear Mixed Model with Random Intercept and Slope using Unstructured Covariance Matrix between Intercept and Slope Estimates; baseline UPDRS; and Center adjusted.

### Hypothesis #1: Supportive/Sensitivity Analyses

The Sponsor also compared the slope differences between rasagiline groups vs. placebo based on the completers (CO) and per protocol (PP) samples. Significant effect of each dose level over placebo was shown in the PP analysis set, as well as for 1 mg in the CO analysis set.

As a sensitivity analysis for evaluating Hypothesis#1, the Sponsor used a multiple imputation (MI) approach to replace missing data in the PC phase. The findings are similar to the findings obtained from the primary analysis conducted on the ITT sample. In the ADAGIO study, the comparison between placebo to the 1 mg and 2 mg early-start rasagiline groups were significantly different from zero ( $p\text{-value} \leq 0.0158$ ). In TEMPO, the comparison of placebo to the 1 mg and 2 mg early-start rasagiline groups were not significantly ( $p\text{-value} \geq 0.069$ ) different from zero.

### Hypothesis #2: Primary Analysis

#### ADAGIO Study

Table 6 lists the repeated measures mixed model (MMRM) analysis results for the Active treatment phase data. In ADAGIO, a significant ( $p\text{-value}=0.0250$ ) deterioration from baseline to Week 72 for the 1 mg delayed-start treatment group compared to the 1 mg early-start treatment group was demonstrated. Comparison of change from baseline showed a favorable effect of early treatment with rasagiline 1 mg over delayed (LSMEAN change: -1.680 UPDRS units). However, the difference between the 2 mg early-start and delayed-start groups was not significant ( $p\text{-value}=0.602$ ).

If the combined datasets (1mg early plus 2 mg early and the 1 mg delayed combined with the 2 mg delayed) for the mean change from baseline to Week 72 are analyzed using the primary pre-specified analysis for hypothesis # 2, the p-value increases to 0.0506 (instead of  $p\text{-value}=0.025$ ).

## TEMPO Study

In the TEMPO study, the difference in LSMEAN changes between the 2 mg early-start and delayed-start groups was not significant (p-value=0.076) (Table 6) in the MMRM analysis of the Active treatment phase data.

Table 6. Comparison of Changes from Baseline to Week 72 (ADAGIO)/Week 52 (TEMPO) in Total UPDRS Scores – ACTE Data Analysis Set

Comparison	ADAGIO at Week 72					TEMPO at Week 52				
	Est	SE	P-Value	Lower 95% CI Limit	Upper 95% CI Limit	Est	SE	P-Value	Lower 95% CI Limit	Upper 95% CI Limit
1 mg Early- 1 mg Delayed Start	-1.680	0.747	0.025 <sup>§</sup>	-3.148	-0.212	Not Applicable				
2 mg Early-2 mg Delayed Start	0.356	0.684	0.602	-0.989	1.702	-1.934	1.089	0.076	-4.078	0.209

Source: study reports

MMRM analysis was performed

<sup>§</sup> if the combined dataset ( i.e., including patients belong to 1mg early and delay, and 2mg early and delay) is used for the primary pre-specified analysis for hypothesis # 2, then the p-value is 0.0506.

As a sensitivity analysis for evaluating Hypothesis#2, Sponsor used a multiple imputation (MI) approach to replace missing data in the Active Treatment phase. The findings are similar to the findings obtained from the primary analysis. In ADAGIO, the comparison between the 1 mg delayed-start treatment group vs. the 1 mg early-start treatment group was significant (p-value=0.042, LSMEAN difference=-1.192 at week 72). The comparison between the 2 mg delayed-start treatment group vs. the 2 mg early-start treatment group was not significant (p-value=0.920, LSMEAN difference=0.053 at Week 72). In TEMPO, the comparison between the 2 mg delayed-start treatment group vs. the 2 mg early-start treatment group was also not significant (p-value=0.180, LSMEAN difference=-1.231 at Week 52).

### Hypothesis #3: Primary Analysis

Table 7 lists the hypothesis test results for the noninferiority of slopes during the active-treatment phase with a margin of 0.15 UPDRS points per week. In the ADAGIO study, an identical rate of deterioration in UPDRS was evident in the 1 mg delayed-start and 1 mg early-start treatment groups (each 0.085 UPDRS units/week). The point estimate for the difference between the treatment groups for the change in UPDRS per week was 0.00 with 90% confidence intervals of -0.036 to 0.036. Since the upper limit of the confidence interval did not exceed the predefined upper non-inferiority boundary of 0.15 UPDRS units/week, the null hypothesis was rejected and the 1 mg early start was declared as not inferior to the delayed-start, with regard to deterioration rates in the active phase.

The deterioration in UPDRS was also seen in the 2 mg delayed-start and 2 mg early-start treatment groups (0.065 and 0.094 UPDRS units/week, respectively). The point estimate for the difference between the treatment groups in the change in UPDRS per week was 0.029 with 90% confidence intervals of -0.005 to 0.062. The upper confidence interval did not exceed the predefined upper boundary of 0.15 UPDRS units/week.

According to the hierarchical method of adjustment for multiple comparisons, the 2 mg rasagiline groups should not be tested on the hypothesis #3, and therefore no conclusion is made regarding the statistical significance of this comparison for both ADAGIO and TEMPO studies.

Table 7. Comparison of Changes per Week (Slopes) in Total UPDRS during Active-Treatment Phase (ADAGIO Weeks 48-72/TEMPO Weeks 42-52) - ACTE Data Analysis Set

Comparison	ADAGIO				TEMPO			
	Est	SE	Lower 90% CI Limit	Upper 90% CI Limit	Est	SE	Lower 90% CI Limit	Upper 90% CI Limit
1 mg Early-1 mg Delayed Start Slope Difference	0.000	0.022	-0.036	0.036	N/A			
2 mg Early-2 mg Delayed Start Slope Difference	0.029	0.020	-0.005	0.062	-0.100	0.081	-0.234	0.033

Linear Mixed Model with Random Intercept and Slope using Unstructured Covariance Matrix between Intercept and Slope Estimates  
Separate analysis for each dose level ( in ADAGIO study)

As a sensitivity analysis for evaluating Hypothesis#3, the Sponsor used a multiple imputation (MI) approach to imputing missing data in the Active Treatment Phase, the findings are similar to the findings obtained from the primary analysis as stated in Table 7.

Comments from Agency Statistical Reviewer regarding the protocol specified primary analysis.

The Agency’s Statistician was able to reproduce the Sponsor’s reported efficacy findings for Hypothesis#1 in the PC phase in both studies. Although it was assumed that the full symptomatic effect of rasagiline would be present before week 12 the analysis of the observed data does not support this assumption. The data supported a conclusion of a nonlinear relationship of the data between weeks 12 and 36, hence the symptomatic benefits were not fully established by the end of week 12. The significances of the slope differences in the original analysis were mainly due to the nonlinear relationship of the data between weeks 12 and 36 (Figure 3). The LSMEAN trends of the placebo, 1 mg, and 2 mg rasagiline are listed in Figure 3. The Agency’s statistician reanalyzed the data from ADAGIO excluding the week 12 data from the analyses. Table 8 lists the finding based on including or excluding the 12 week data. In the ADAGIO study, the slope differences between 1mg vs. placebo (p-value=0.098), and 2 mg vs. placebo (p-value=0.834) were not statistically significant. The statistically significant slope differences reported by the Sponsor in the original analysis were mainly due to the nonlinearity among the data points between weeks 12 and 36.

In Figure 4, the LSMEAN trends for the treatment groups in TEMPO study were linear over time and consistent conclusions were drawn based on the inclusion and exclusion of the Week 14 data in the analyses. Therefore, a comparison of the findings obtained from ADAGIO and TEMPO studies supported that it was reasonable to assume that the symptomatic effect associated with rasagiline was not fully evolved by the end of week 12.

In testing Hypothesis#2 at AC phase, this reviewer was able to reproduce the efficacy findings reported for ADAGIO study. Although the rasagiline 1 mg dose was significant in the AC phase, the 2 mg dose of rasagiline failed to demonstrate a disease modifying benefit in the AC phase. The Sponsor did not find a valid explanation for the insignificance of the rasagiline 2 mg dose.

Based on this reviewer’s analysis (excluding week 12 data from the PC phase data analysis), the 1 mg dose of rasagiline failed to demonstrate a disease modifying benefit after multiplicity adjustment.

In TEMPO, the Sponsor included the 1 mg data in the comparison of rasagiline 2 mg early-start vs. delayed-start groups in the AC phase. Since there was no 1 mg delayed-start group in the AC phase, the 1 mg early-start group should not have been included in the analysis for evaluating rasagiline 2 mg early-start vs. delayed-start groups. This reviewer reanalyzed the data for rasagiline 2 mg early-start vs. delayed-start groups excluding the 1 mg data. The LSMEAN difference for rasagiline 2 mg early-start vs. delayed-start groups at week 52 was -1.52 and the p-value for the comparison was 0.133. The TEMPO results do not support a conclusion of efficacy for the 2mg rasagiline dose either in the PC or AC phases.

Table 8: Comparison of Changes per Week (Slope) for Rasagiline vs. Placebo during PC Phase (ADAGIO: Weeks 12, 24, 36; TEMPO: Weeks 14, 20, 26) – Efficacy ITT Data Analysis Set

Comparison	ADAGIO (Weeks 12, 24, 36)			TEMPO (Weeks 14, 20, 26)		
	Est	SE	P-Value	Est	SE	P-Value
1 mg-Placebo Slope Difference	-0.046	0.019	0.0133	-0.085	0.057	0.1342
2 mg-Placebo Slope Difference	-0.072	0.019	0.0001	-0.083	0.057	0.1475
	ADAGIO (Weeks 24, 36)			TEMPO (Weeks 20, 26)		
1 mg-Placebo Slope Difference	0.049	0.029	0.098	-0.098	0.101	0.372
2 mg-Placebo Slope Difference	-0.006	0.029	0.834	-0.049	0.111	0.662

Est: Estimate

Repeated Measures Mixed Linear Model with Random Intercept and Slope using Unstructured Covariance Matrix between Intercept and Slope Estimates; Baseline UPDRS and Center adjusted.

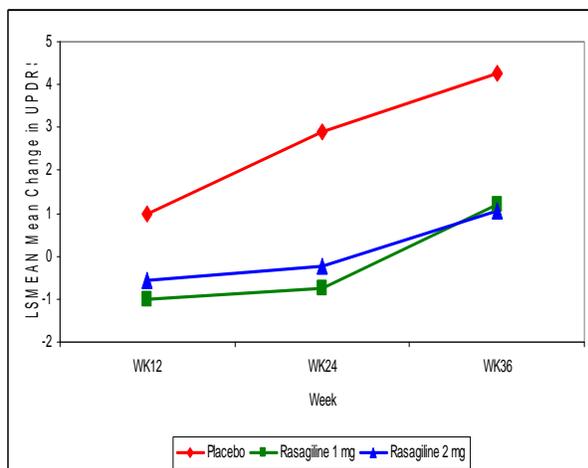


Figure 3: ADAGIO STUDY

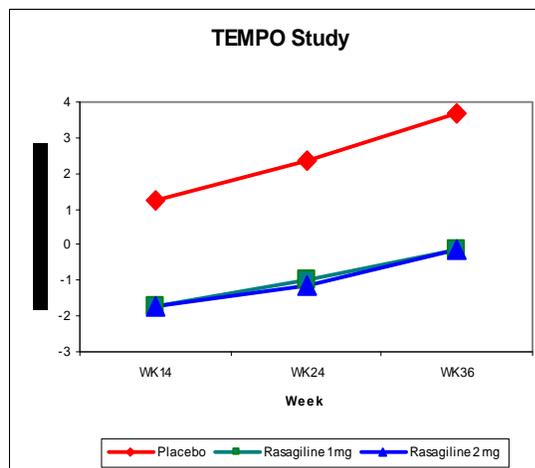


Figure 4. TEMPO STUDY

The significance of Hypothesis#2 in the AC phase of ADAGIO seems to be dependent on whether the combined (subjects belonging to the 1mg plus 2 mg early and the combined 1 and 2 mg delayed groups) or separate (only the subjects who belong to the 1mg early and delay groups) are compared in the MMRM model. In the statistical analysis plan, the Sponsor did not pre-specify whether combined or separate data sets would be analyzed in evaluating the primary hypotheses. When performing an analysis of the combined data sets from a multi-doses study, it is assumed that the variability of the outcome measure across the doses tested are same, the model design assumes there is no interaction between dose level and the pre-specified covariates. If this assumption is not valid for a data set, then a separate model for comparison of each dose vs. placebo is required to allow for valid statistical inference. The ADAGIO (see Table 9) study, the estimated standard errors (SE) for the comparisons of (i) 1 mg Early vs. 1 mg Delayed Start, and (ii) 2 mg Early vs. 2 mg Delayed Start obtained from separate analyses were 0.742 and 0.684, respectively. Whereas the corresponding SEs obtained from combined analysis were 0.727 and 0.712, respectively. The above findings were also supported by the significances of the two interaction terms: *dose level by baseline UPDRS* (p-value=0.048) and *dose level by center* (p-value=0.012) in the model. Since the two comparisons had different SEs, it was justifiable to consider separate model data analysis for the two comparisons. The LSMEAM estimates were also different in the separate and combined analysis. Therefore, the separate model data analysis is appropriate for testing the hypothesis#2.

Table 9: A Comparison of the Findings from a Combined Data vs. Separate Data Analysis- ADAGIO Study

Comparisons (evaluated from separate model analysis)	ADAGIO at Week 72		
	Est	SE	P-Value
1 mg Early- 1 mg Delayed Start	-1.680	0.747	0.025
2 mg Early-2 mg Delayed Start	0.356	0.684	0.602
Comparisons (evaluated from a combined model analysis)			
1 mg Early- 1 mg Delayed Start	-1.42	0.727	0.050
2 mg Early-2 mg Delayed Start	0.179	0.712	0.801

The Sponsor submitted a post hoc analysis for Hypothesis #2 by comparing the effects of baseline UPDRS scores divided into quartiles ( $\leq 14$ ,  $>14-\leq 19$ ,  $>19-\leq 25.5$ ,  $>25.5$ ) on the ADAGIO study data (see Table 10). The stated purpose was to demonstrate that there is an increased benefit associated with early vs. delayed treatment when variations in disease severity are considered. However, the LSMEAN difference estimates between early vs. delayed did not support such a claim. For example, for the 1 mg dose, the LSMEAN difference was -1.851 in the second quartile (Baseline UPDRS  $>14-\leq 19$ ), whereas, the difference was -0.0177 for the third quartile (Baseline UPDRS  $>19-\leq 25.5$ ). In the TEMPO trial data, there was also no evidence of an increased benefit of early vs. delayed treatment with increased disease severity.

Since both the 1mg and 2mg rasagiline groups fail to demonstrate sufficient evidence of a DM benefit (based on total UPDRS scale score) in ADAGIO, there is no interest in looking for efficacy by dose for the individual subparts (I, II, and III) of the UPDRS. As a result, there is no statistical review of individual subpart of the UPDRS scale.

Table 10. Hypothesis #2- Differences in the Changes in Total UPDRS from Baseline to Week 72 by Baseline UPDRS subgroups and Dose Level

ADAGIO Study			Early-Delayed Difference	
Variable	Dose Level	Subgroups	LSMEAN difference Estimate	p-value
Baseline UPDRS Quartiles in ADAGIO	1 mg Rasagiline	Baseline UPDRS $\leq 14$	-1.0965	0.4579
		Baseline UPDRS $>14-\leq 19$	-1.8510	0.2162
		Baseline UPDRS $>19-\leq 25.5$	-0.01779	0.9909
		Baseline UPDRS $>25.5$	-4.0197	0.0152
	2 mg Rasagiline	Baseline UPDRS $\leq 14$	0.8884	0.5145
		Baseline UPDRS $>14-\leq 19$	0.7879	0.5599
		Baseline UPDRS $>19-\leq 25.5$	1.5629	0.2746
		Baseline UPDRS $>25.5$	-2.4693	0.1036
TEMPO				
Variable	Dose Level	Subgroups	Estimate	p-value
Baseline UPDRS Quartiles in TEMPO	2 mg Rasagiline early vs. delayed	Baseline UPDRS $\leq 17.5$	-3.153	0.0842
		Baseline UPDRS $>17.5-\square 23$	-3.711	0.0616
		Baseline UPDRS $>23-\square 23.5$	4.476	0.0398
		Baseline UPDRS $>31.5$	-6.050	0.0284
Baseline UPDRS Quartiles in ADAGIO	2 mg Rasagiline early vs. delayed	Baseline UPDRS $\leq 14$	-3.433	0.1851
		Baseline UPDRS $>14-\leq 19$	-3.066	0.1374
		Baseline UPDRS $>19-\square 25.5$	0.047	0.9803
		Baseline UPDRS $>25.5$	-1.713	0.3822

Source: ISE report

## NATURAL HISTORY STAGGERED START (NHSS) METHOD

The Sponsor submitted results using a new analysis method called natural history staggered start analysis (NHSS) method as a supportive analysis to estimate Disease Modifying (DM) effect. The NHSS analysis model is defined as follows:

Let  $x$  be an indicator of treatment assignment ( $x = 1$  if the patient is assigned to treatment,  $x = 0$  if assigned to placebo) and let  $y_0$  be the centered baseline value of the clinical outcome (difference between a patient's baseline value and the overall baseline mean). The model for the mean change from baseline in the clinical outcome at post-randomization time  $k$ , denoted  $t_k$  and measured in years for simplicity, can be written as a simple linear model:

$$\mu_k = \alpha_0 + \alpha_1 x + \tau_0 y_0 + \tau_1 x y_0 + \beta_0 t_k + \beta_1 x t_k + \gamma_0 y_0 t_k + \gamma_1 x y_0 t_k,$$

and the model for the change from baseline for an individual can be written as:

$$\Delta y_{ik} = \alpha_0 + \alpha_1 x_i + \tau_0 y_{i0} + \tau_1 x_i y_{i0} + \beta_0 t_k + \beta_1 x_i t_k + \gamma_0 y_{i0} t_k + \gamma_1 x_i y_{i0} t_k + e_{ik}$$

where  $\alpha_1$  and  $\tau_1$  correspond to symptomatic effects, and  $\beta_1$  and  $\gamma_1$  correspond to disease modifying effects. The parameter  $\beta_0$  is the slope of the placebo group and  $\beta_1$  is the difference in the slopes of the treatment and placebo groups for a patient of average severity at baseline. The parameter  $\gamma_0$  is the change in the slope of the placebo group associated with a one point increase in severity at baseline, and  $\gamma_1$  is the change in the difference in slopes between the treatment and placebo groups associated with a one point increase in severity at baseline. A Natural History Estimator (NHE) is defined as

$$NHE = \frac{\beta_1 - \beta_0 \tau_1}{1 + \tau_1}$$

The standard error of the NHE is estimated using a bootstrapping procedure that re-sampled from the original data set with replacement. The NHE can be interpreted as the number of points per year of treatment-related benefit (relative to placebo) due to disease modification, or specifically based on the proposed definition of disease modification, the treatment benefit that is accumulating over time.

### NHSS Results:

The Sponsor reported NHSS results for the following four Models.

Model 1: "Placebo vs. Early Treatment" includes data from the first phase of the study plus continuing data from the second phase for patients who started treatment early but excludes second phase data for delayed start patients. In fitting Model 1 for ADAGIO study, the data collected at Baseline, weeks: 12, 24, & 36 for placebo group, and data collected at Baseline, weeks 12, 24, 36, 42, 48 54, 60, 66, & 72 were used for Early Treatment group.

Similarly, for TEMPO, the data collected at Baseline, weeks: 12, 20, and 24 for placebo group, and data collected at Baseline, weeks 12, 20, 24, 42, 54, and 60 were used for the Early Treatment group in Model 1.

Model 2: “Placebo vs. Pooled” includes data from the first phase of the study plus data from the second phase for patients who were in the delayed start group. Data from the second phase for the delayed start patients was “pooled” with the data from the first phase by “moving the data back” to time 0, so that time represented the time on the current treatment. Also, the baseline values were set to the value at the time that treatment was started. This model includes second phase data for early start patients.

Model 3: “Placebo vs. All” includes all data from Models 1 and 2.

Model 4: “First Phase Only” includes only data from the placebo controlled phase of the study.

Table 11. Disease Modification Estimates (ADAGIO and TEMP studies)

ADAGIO	Group	Model	$\beta_0$	$\beta_1$	$\tau_1$	% Reduction in Decline from Placebo rate	NHE (points per year)	95% Confidence Interval for NHE	
<b>PBO vs. Early</b>	<b>1 mg</b>	Model: 1	7.39	-3.38	-0.06(n.s)	<b>42.3%</b>	<b>-3.13*</b>	<b>-5.06</b>	<b>-1.19</b>
PBO vs. Pooled	1 mg	Model: 2	7.68	-3.41	-0.04 (n.s)	42.3%	-3.25*	-5.35	-1.15
PBO vs. All	1 mg	Model: 3	7.40	-3.50	-0.12 (n.s)	39.8%	-2.95*	-4.76	-1.13
First Phase	1 mg	Model: 4	7.42	-2.25	-0.10 (n.s)	22.4%	-1.66	-3.85	0.53
<b>PBO vs. Early</b>	<b>2 mg</b>	<b>Model: 1</b>	7.48	-4.27	-0.01	<b>56.8%</b>	<b>-4.25*</b>	<b>-5.72</b>	<b>-2.77</b>
PBO vs. Pooled	2 mg	Model: 2	7.57	-4.24	-0.07	52.9%	-4.00*	-6.02	-1.97
PBO vs. All	2 mg	Model: 3	7.48	-4.06	-0.15	46.0%	-3.44*	-5.09	-1.79
First Phase	2 mg	Model: 4	7.47	-4.13	-0.01	55.1%	-4.11*	-6.24	-1.98
<b>TEMPO</b>									
First Phase Only	2 mg	Model: 4	10.60	-4.20	-0.06	36.0%	-3.81	-11.2	3.59

Source; ISE report

\* Statistically significant at significant level 0.05

Table 11 lists the Sponsor’s reported NHSS results for the ADAGIO and TEMPO studies. In ADAGIO study, according to the reported findings in the Table 9, Model 1, Model 2, & Model 3 demonstrated a statistically significant DM effect. Models 1, 2, & 3 demonstrated DM effects of -3.13, -3.25, and -2.95 UPDRS points/year, respectively. Model 4 (analysis of data from the first phase only) did not demonstrate a statistically significant DM effect for the 1mg compared to placebo. In the TEMPO study, Model 4 failed to show a significant DM effect for the 2mg early compared to Placebo.

Comments from Agency's Statistician Regarding NHSS Modeling and the Findings of a DM Effect.

In the first group of subjects (Model 1 in Table 11), the data collected at Baseline, plus weeks: 12, 24, & 36 for the placebo group, and data collected at Baseline, plus weeks 12, 24, 36, 42, 48 54, 60, 66, & 72 for the 1mg Early were used in analysis. That is, data from only three post-baseline visits were included for subjects in the placebo group in the NHSS model, whereas data from nine post- baseline visits were included in the model for the 1mg Early Treatment group. The reliability of statistical inferences based on such inequality of data points between two groups seems to be always questionable or an unrealistic statistical practice. Therefore, the findings based on model 1 are not acceptable for consideration as supportive evidence for a disease modifying effect. For the same reason, the findings obtained from the analysis using Models 2 and 3 are not acceptable. In both cases (Model 2 and Model 3), there was the same inequality in the number of data points between the placebo and treatment groups. Along with some other clarification questions, the Sponsor was also asked to provide statistical justification for using NHE analysis on models 1, 2, and 3 (request date May 12, 2011. The Sponsor did not respond on this question in responding other questions (response date July 29, 2011).

In Model 4, the data obtained at PC Phase for the placebo and the 1 mg Early drug groups were analyzed in the NHSS model. The estimated NHE (-1.66 per year) is not statistically significant (95% CI: -3.85, 0.53). Therefore, there is no statistical evidence of disease modifying effect for 1 mg Early start drug group.

The NHE was defined as the  $NHE = \frac{\beta_1 - \beta_0 \tau_1}{1 + \tau_1}$ . If  $\tau_1$  (the slope of baseline\*treatment) is not statistically significant for a given data set (i.e., the estimated  $\tau_1$  is not different from zero), then the estimated NHE tends to be  $\beta_1$ . That is, the NHE is comparable to the slope difference between 1mg Early and placebo groups at PC Phase. In this data set,  $\tau_1$  was not statistically significant (p-value=0.0786). The observed mean UPDRS total scores at baseline by treatment groups also supported the conclusion that baseline scores were not different across treatment groups (Table 12). The insignificance of  $\tau_1$  indicates that there was no change in the difference in slopes between the 1mg Early and placebo groups associated with an increase/decrease in severity of the disease at baseline. Similarly, in the 2mg Early vs. placebo groups NHSS analysis (Model 4),  $\tau_1$  was also not statistically significant (p-value=0.922). Similarly, for the 2mg Early vs. Placebo groups in TEMPO data, the estimated  $\tau_1$  (the slope of baseline\*treatment) was not statistically significant (p-value=0.639).

The insignificance of  $\tau_1$  supports the belief that the  $NHE = \beta_1$ . That is, the NHSS analysis (for Model 4) addresses the model (a Linear Mixed Model with random intercept and slope) which was used for testing the protocol specified Hypothesis #1. Extensive discussions over the last few years involving researchers from the pharmaceutical industry, academia, and the FDA resulted in a belief that a comparison limited to only the slopes between study drug and placebo in the PC phase was not sufficient evidence to support a claim for a disease modifying

effect of associated with a drug. An additional sequence of hypothesis testing (Hypotheses#2 & 3 as specified in the protocol) were required to confirm a disease modifying effect of a drug. Hence, the proposed NHSS analysis does not appear to be an appropriate approach for evaluating a drug for a potential disease modifying effects.

Table 12. Mean scores of Total UPDRS Score at Baseline by Treatment group

TVP-1012/500 (ADAGIO)		1 mg Delayed Start	1 mg Early Start	2 mg Delayed Start	2 mg Early Start
Baseline UPDRS Total	N	298	288	295	293
	Mean	20.2	20.6	19.9	20.8
TVP-1012/232 (TEMPO)		1 mg	2 mg	Placebo	
Baseline UPDRS Total	N	134	132	138	
	Mean	24.69	25.89	24.54	

Source: Study Reports

As requested by the Agency, the Sponsor submitted the results of a simulation study for the NHSS model. In the simulation study result, the Sponsor generated baseline scores independently from the post-baseline scores. In generating the post-baseline scores across three visits, an autoregressive process (AR(1)) was used. In clinical trials, the post-baseline scores are highly correlated with the patient’s baseline score. In addition, in clinical trials, there is no evidence of a presence of AR(1) process in the repeated measure data within an individual patient. Generally, an unstructured covariance/correlation (UN) is observed within the repeated measures within an individual patient in clinical trials. The ADAGIO study data also support the presence of an UN covariance structure in the repeated measures within an individual patient. Moreover, in the original data analyses of the ADAGIO study, an UN covariance was used in both protocol specified primary and NHSS analyses. So, a consideration of an AR(1) process is not acceptable in the simulation study. In a teleconference, the Sponsor was informed the Agency’s concerns of (i) generating baseline score independent from the post-baseline scores and (ii) the use of AR(1) process. In the same teleconference, the Sponsor was requested to submit another simulation study result considering an UN covariance structure among the baseline and post-baselines scores. However, the Sponsor generated again the baseline score independent from the post-baseline scores and then introduced a correlation with the post-baseline scores. The included correlation between baseline score and post-baseline scores was relatively weaker as compared to the corresponding observed correlation in the ADAGIO study data set. Since the baseline scores seems to be the key factors in NHSS analysis, the Sponsor needs to generate data from a multivariate distribution considering the obtained covariance structure from the ADAGIO study data points.

The simulation study under a null (i.e., NHE=0) in NHSS analysis is nothing but a simulation study under a null in a linear mixed model (i.e., a slope difference between two groups=0 [ $\beta_1 = 0$  in NHSS model;  $NHE = \frac{\beta_1 - \beta_0 \tau_1}{1 + \tau_1}$ ]). The NHE can have zero (or close to zero) only when

$\beta_1 = 0$ . In the presence of a curved relationship over the data points, both the linear mixed model and NHSS model analyses inflate Type I error rate. In the ADAGIO study, the data in Phase I form a curved relationship among the three data points for both treatment groups (1 mg, and 2 mg rasagiline) is observed (see Figure 3), hence both the linear mixed model and NHSS model analysis inflate Type I error rate, and the resulting p-values obtained from the two models are questionable. Therefore, the submitted simulation study results are not useful for evaluating the statistical properties (e.g., Type I) of a NHSS model.

## 4. SUBGROUP ANALYSIS

### Subgroup Analyses in ADAGIO and TEMPO studies

Three primary hypotheses were evaluated in a post-hoc manner looking for a relationship with demographic characteristics: Sex, Age by median ( $\leq 63$ ,  $>63$  years), and North America vs. Rest of the World (ROW). The findings did not differ systematically across subgroups within each study and they did not support an indication of a disease modifying effect of rasagiline in the sequence of testing the three hypotheses for any of the subgroups (Table 13 and Table 14).

Table 13. Subgroup Analysis on the Primary Efficacy Measure UPDRS Total Score-ADAGIO Study.

ADAGIO Study					
Variable		Dose Level	Subgroups	Slope Estimate	Standard Error (SE)
Hyp #1	Gender	1 mg Rasagiline	Female	0.063	0.021
			Male	0.108	0.016
		2 mg Rasagiline	Female	0.041	0.018
			Male	0.083	0.017
	Baseline Median Age	1 mg Rasagiline	Age $\leq 63$	0.092	0.019
			Age $> 63$	0.092	0.018
		2 mg Rasagiline	Age $\leq 63$	0.063	0.016
			Age $> 63$	0.071	0.019
	Geographical Area	1 mg Rasagiline	North America	0.107	0.023
			Rest of the World	0.078	0.015
		2 mg Rasagiline	North America	0.104	0.021
			Rest of the World	0.039	0.016
				Early group -Delayed group( LSMEAN Difference at week 72)	
Variable		Dose Level	Subgroups	Estimate	SE
Hyp #2	Gender	1 mg Rasagiline	Female	-3.858	1.364
			Male	-0.046	1.097
		2 mg Rasagiline	Female	-1.755	1.314
			Male	1.025	0.930
	Baseline Median	1 mg Rasagiline	Age $\leq 63$	-1.752	1.365
			Age $> 63$	-1.846	1.108

	Age	2 mg Rasagiline	Age ≤ 63	-0.079	1.244
			Age > 63	0.244	0.953
	Geographical Area	1 mg Rasagiline	North America	-1.642	1.341
			Rest of the World	-1.649	0.977
		2 mg Rasagiline	North America	1.027	1.150
			Rest of the World	-0.253	0.941
				Early group - Delayed group (Diff. in slope)	
Variable		Dose Level	Subgroups	Estimate	SE
Hyp #3	Gender	1 mg Rasagiline	Female	-0.062	0.026
			Male	0.039	0.020
		2 mg Rasagiline	Female	0.022	0.024
			Male	0.037	0.019
	Baseline Median Age	1 mg Rasagiline	Age ≤ 63	-0.038	0.025
			Age > 63	Model does not converge	
		2 mg Rasagiline	Age ≤ 63	0.023	0.022
			Age > 63	0.038	0.020
	Geographical Area	1 mg Rasagiline	North America	0.023	0.027
			Rest of the World	-0.019	0.196
2 mg Rasagiline		North America	0.047	0.026	
		Rest of the World	0.019	0.017	

Source: ISE Report

Table 14. Subgroup Analysis on the Primary Efficacy Measure UPDRS Total Score-TEMPO Study.

TEMPO Study					
Variable		Dose Level	Subgroups	Estimate (Slope)	SE
Hyp #1	Gender	2 mg Rasagiline	Female	0.180	0.058
			Male	0.104	0.051
	Baseline Median Age	2 mg Rasagiline	Age ≤ 63	0.142	0.053
			Age > 63	0.118	0.056
				Early group - Delayed group (LSMEAN Difference Week 52)	
Variable		Dose Level	Subgroups	Estimate	SE
Hyp #2	Gender	2 mg Rasagiline	Female	-1.875	2.198
			Male	-2.355	1.515
	Baseline Median Age	2 mg Rasagiline	Age ≤ 63	-1.353	1.458
			Age > 63	-1.733	1.658
				Early group-Delayed group (Slope Diff)	
Variable		Dose Level	Subgroups	Estimate	SE
Hyp #3	Gender	2 mg Rasagiline	Female	0.018	0.264
			Male	0.012	0.015

	Baseline Median Age	2 mg Rasagiline	Age ≤ 63	0.028	0.016
			Age > 63	-0.007	0.017

Source: ISE Report

## 5. SUMMARY AND CONCLUSIONS

Based on the findings of the protocol specified statistical analysis methods in the ADAGIO study, (i) the null hypothesis (for hypothesis #1) was rejected for the comparison of placebo to 1 mg rasagiline early-start and the placebo to 2 mg early-start comparison; (ii) the null hypothesis (for hypothesis #2, for the 1 mg rasagiline delayed-start to early-start comparison) was rejected and the null hypothesis for the 2 mg rasagiline delayed-start to early-start comparison was not rejected; and (iii) the null hypothesis (for hypothesis #3) for the 1 mg early to 1 mg delayed-start comparison was rejected. According to the hierarchical method of adjustment for multiple comparisons, the 2 mg rasagiline groups was not tested on hypothesis #3, therefore no conclusion was made regarding the statistical significance of this comparison.

In the reviewer's analyses for the hypothesis #1 (excluding week 12 data from the PC phase data analysis), the null hypotheses (for hypothesis #1) for both rasagiline 1 mg and 2 mg doses were not rejected, hence hypothesis #2 and hypothesis #3 should not be formally tested according to the hierarchical method of adjustment for multiple comparisons. That is, the results of the ADAGIO study fail to demonstrate a disease modifying benefit associated with the 1 mg or 2 mg dose of rasagiline.

The analyses conducted by the Sponsor and this reviewer conclude that the TEMPO study data did not support a disease modifying benefit for the rasagiline 2mg dose based the protocol specified statistical analysis methods used for the primary analysis of the ADAGIO study data.

As supportive evidence, the Sponsor submitted NHSS approach analysis. This reviewer finds that the slopes ( $\tau_1$ ) for baseline\*treatment were not statistically significantly different from zero in both ADAGIO and TEMPO studies. The insignificance of  $\tau_1$  supports that belief that the NHE= $\beta_1$ . That is, the NHSS analysis (for Model 4) is comparable to the model (a Linear Mixed Model with random intercept and slope) which was used for testing Hypothesis #1. In discussions between agency and the Sponsor over the last few years, it was understood that slope comparison between the study drug group and the placebo group in the PC phase would not provide sufficient evidence to support a claim for a disease modifying effect for rasagiline. An additional sequence of hypothesis testing (Hypotheses#2 & 3) is required to confirm a disease modifying effect. Hence, the proposed NHSS analysis does not seem to be an appropriate approach for evaluating drugs for a potential disease modifying effect.

The Sponsor submitted sensitivity analyses /secondary analyses for dealing with the missing data in the two studies. The sensitivity analyses do not support the presence of any disease modifying benefit of the study drug or the methods are not appropriate for evaluating disease modifying benefit.

The subgroup analyses (considering the sequence of hyp#1, hyp#2 & hyp#3) for the two studies also do not provide evidence for a disease modifying claim in any of the subgroups.

Although the Sponsor claims that the data for the 1 mg dose of rasagiline (in ADAGIO) and the 2 mg dose (in TEMPO) demonstrate a disease modifying benefit in early untreated patients with idiopathic Parkinson's disease, this reviewer's analysis of the study data does not support the conclusion of a disease modifying benefit associated with either dose of rasagiline.