Made available by Hasselt University Library in https://documentserver.uhasselt.be

Assessment of type I error rate associated with dose group switching in a longitudinal Alzheimer trial

Non Peer-reviewed author version

HABTEAB GHEBRETINSAE, Aklilu; MOLENBERGHS, Geert; Dmitrienko, Alex; Offen, Walt & Sethuraman, Gopalan (2014) Assessment of type I error rate associated with dose group switching in a longitudinal Alzheimer trial. In: Journal of Biopharmaceutical Statistics, 24 (3), p. 660-684.

DOI: 10.1080/10543406.2014.888570
Handle: http://hdl.handle.net/1942/18607

Assessment of Type I Error Rate Associated with Dose Group Switching in a Longitudinal Alzheimer Trial

Aklilu Habteab Ghebretinsae[1] **Geert Molenberghs**[1,2]

Alex Dmitrienko[3] **Walt Offen**[4] **Gopalan Sethuraman**[4]

[1] *I-BioStat, Universiteit Hasselt, B-3590 Diepenbeek, Belgium*

[2] *I-BioStat, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium*

[3] *Quintiles Innovation, Overland Park, Kansas, USA*

[4] *Eli Lilly & Company, Lilly Corporate Center, Indianapolis, IN*

Abstract

In clinical trials, there always is the possibility to use data-driven adaptation at the end of a study. There prevails, however, concern on whether the type I error rate of the trial could be inflated with such design, thus necessitating multiplicity adjustment. In this project, a simulation experiment was set up to assess type I error rate inflation associated with switching dose group as a function of drop out rate at the end of the study, where the primary analysis is in terms of a longitudinal outcome. This simulation is inspired by a clinical trial in Alzheimer's disease. The type I error rate was assessed under a number of scenarios, in terms of differing correlations between efficacy and tolerance, different missingness mechanisms, and different probabilities of switching. A collection of parameter values was used to assess sensitivity of the analysis. Results from ignorable likelihood analysis show that the type I error rate with and without switching, was approximately the posited error rate for the various scenarios. Under LOCF, the type I error rate blew up both with and without switching. The type I error inflation is clearly connected to the criterion used for switching. While in general switching, in a way related to the primary endpoint, may impact the Type I error, this was not the case **for most scenarios in** the longitudinal Alzheimer trial setting under consideration, where patients are expected to worsen over time.

Some Keywords: LOCF; Ignorable likelihood; MAR; MNAR; Type I error.

1 Introduction

In clinical trials, it is not uncommon to modify trial and/or statistical procedures during conduct, based on review of interim data, or even at the end of the study. Such adaptive designs have been in use for quite a while now. Procedural changes may also be implemented at the end of the study; this is of interest here. Adaptation oftentimes reflects medical practice, and may be regarded as ethical

conduct with respect to both efficacy and tolerance of the experimental treatment. However, it is a concern whether the p-value or the confidence interval associated with the treatment effect obtained modification can be reliably and correctly be compared to the nominal α level (Chang 2008, Chow and Chang 2008, Wang, Wu, and Tsai 2008).

The objective of this paper is to examine key operating characteristics of a clinical trial design with data-driven adaptation, when the primary analysis is based on a longitudinal outcome. The assessment of the type I error rate inflation associated with adaptation of a trial by switching dose groups at the end of the study is scrutinized in particular. This work is motivated by a clinical trial in Alzheimer's disease. It is believed by some that switching dose groups may lead to an inflated Type I error rate and thus the significance level needs to be adjusted. On the other hand, it may not be applicable to some trials, including the Alzheimer's disease trial considered here. The Type I error rate can be lower for such trials, thus not inflating the overall Type I error rate, when the primary analysis is modified by switching doses based on high dropout rate in the high dose arm.

Here, the Type I error rate inflation related with switching the dose level in the primary analysis is assessed by comparing the estimated Type I error rate with switching doses and without switching, i.e., adhering to the pre-specified comparison.

The remainder of the manuscript is organized as follows. Section 2 offers some theoretical background on the inflation of Type I error rate associated with switching treatment comparison. Section 3 describes the design of the Alzheimer clinical trial considered and the statistical model employed. Section 4 presents the design of the simulation study, the results of which are described in Section 5.

2 Theoretical Background on Changes in the Primary Analysis and Their Effects on Type I Error Rate

We consider a clinical trial with a primary endpoint of which the primary analysis can be performed using two different comparisons. A priori, the intention is to use comparison 1 (High dose *versus* Placebo), but one can switch to comparison 2 (Low dose *versus* Placebo). The comparison is selected at the end of the study and the decision for switching is driven by the data collected in the study. That is, depending on the drop-out rate in the high-dose arm one of the possible comparisons is

chosen.

Whether switching a comparison is allowed or not (in the latter case, one sticks to the first comparison), the primary analysis is performed at a pre-specified α level. Evidently, a key question is whether this switching strategy inflates the Type I error rate. To address this question, denote D as the test statistic used in the decision for switching rule. If $D < d$, the primary analysis will be performed using the first comparison, otherwise the second comparison is employed. In addition, let t_1 and c_1 represent the test statistic and critical value for comparison 1, respectively. In this case, if $t_1 \geq c_1$, the primary analysis becomes significant, based on comparison 1. Note that, under the hypothesis of no treatment difference, the error rate is approximately the nominal rate [$P(t_1 \geq c_1) \leq \alpha$]. Similarly, take t_2 and c_2 to be the test statistic and critical value for comparison 2. Then, as before, the error rate is approximately the nominal rate [$P(t_2 \geq c_2) \leq \alpha$], under the hypothesis of no treatment difference.

If the trial does not allow for switching, then the comparison is entirely based on the high dose and placebo arms. Evidently, then there is neither a multiple testing problem nor error rate inflation. If switching is allowed, then the type I error rate is given by:

$$[P(D < d)P(t_1 \geq c_1 | D < d)] + [P(D \geq d)P(t_2 \geq c_2 | D \geq d)].$$

Because low dose is not part of the switching criterion, the test of comparison 2 does not depend on the switching process, i.e., t_2 and D are independent. Controlling the type I error rate is equivalent to having $P(t_1 \geq c_1 | D < d) \leq \alpha$. If the decision for switching and testing high dose *versus* placebo comparison are independent then the type I is protected, i.e., $P(t_1 \geq c_1 | D < d) = P(t_1 \geq c_1) \leq \alpha$. However, in general, the Type I error rate may not be preserved at the α level. The amount of Type I error rate inflation is likely to increase with the increase in correlation between the decision for switching and first comparison. This will be scrutinized in the current manuscript.

3 A Clinical Trial in Alzheimer's Disease

Alzheimer's disease (AD) is a neurodegenerative disorder causing progressive decline in memory and other aspects of cognition. The average duration from onset of symptoms to nursing home placement is 5 to 7 years and from symptom onset to death is 7 to 9 years (Figure 1).

Figure 1, About Here.

Although 6-month trials are still standard in regulatory guidelines for AD trials, 18 month long randomized placebo controlled trials are very common. Discontinuation rates for any cause, including death, vary across long-term AD trials and range between 20% and 40%. Thus, it is critical to account for the high discontinuation rates in the design and analysis of AD studies.

A phase III clinical trial was designed for patients who were at least 55 years old to assess the effect of an experimental treatment (ET) compared to placebo on AD progression using co-primary endpoints that include both cognitive and functional measures. This study was a multi-site (176 sites), randomized, double-blind, placebo-controlled, Phase III study of 1500 patients to compare 3 treatments: (i) high dose of the experimental treatment, (ii) low dose of the experimental treatment, and (iii) placebo. The co-primary endpoints for cognition and function were assessed on all patients at baseline (prior to start of treatment) and six post-baseline visits (at weeks 12, 28, 40, 52, 64, and 76). Patients who were on a stable dose of concomitant symptomatic medications (AChEI or memantine) were allowed to stay for the duration of the study.

Patients with mild to moderate AD who met entry criteria were randomized in a 1:1:1 ratio (500 per treatment arm) to 1 of the 3 treatment groups. Patients were randomized by site and severity of AD; mild or moderate AD based on the score from the Mini-Mental State Examination (MMSE) scale.

The primary hypothesis being tested is that the high dose of the experimental drug (HD) slows down the decline rate associated with AD as compared with placebo after 76 weeks of treatment. In other words, the decline for the experimental drug is smaller than the decline for placebo. This can be formulated as:

$$\begin{aligned} H_0 &: \mu_{76,\text{HD}} = \mu_{76,\text{Placebo}}, \\ H_1 &: \mu_{76,\text{HD}} < \mu_{76,\text{Placebo}}, \end{aligned}$$

where μ_{76} is the mean change in decline from baseline. Given that the expected discontinuation rates for the high dose group were unknown, a contingency plan was included in the protocol. This plan stated that if a large proportion of patients in the high dose group dropped out of the study,

then primary comparison would be between the low dose group and placebo. The specific criteria that would trigger the switch were pre-specified as follows: If (1) the discontinuation rate in the high dose group was greater than 50%; (2) the discontinuation rate in the high dose group exceeded the discontinuation rate in the placebo group by 20%; (3) the low dose group does not meet either of the criteria above (1) or (2).

In analyzing the treatment effect, a linear mixed-effects model is considered (Verbeke and Molenberghs 2000), allowing for a direct-likelihood approach to incomplete data, which is sometimes referred to as 'repeated measures mixed model' (MMRM), and is used for analysis of the primary endpoint. The general form of such a model is: $\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i$, where $\boldsymbol{X}_i$ is a design matrix for the fixed effects $\boldsymbol{\beta}$ for subject $i = 1, \ldots, N$, $\boldsymbol{Z}_i$ is a design matrix for the random effects $\boldsymbol{b}_i \sim N(\mathbf{0}, D)$, and $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma}_i)$ are the within-patient random errors.

In our case, the model for the fixed effect includes 8 independent variables: baseline score, age at baseline, treatment, visit (post-baseline assessments; a categorical variable), treatment by visit interaction, MMSE stratification factor at baseline (mild or moderate), concomitant ACHEI or memantine use at baseline (yes or no), and investigator (site). In the final analysis, models with site as a random effect and a model without site are considered. Thus, all in all, we have four cases, referring to site as: (1) fixed effect; (2) random effect; (3) included while generating data but excluded in the data analysis; (4) excluded while generating data as well as in the data analysis.

4 Simulation Study

4.1 Generating Datasets

In line with the Alzheimer trial, our simulation starts from 176 sites. Each site is assumed to have and equal sample size of 9, that is, 3 patients per treatment arm. The total population sums up to $528 \times 3 = 1584$ patients. Every time, a complete dataset is generated. First, for each patient, the values of the covariates are generated from a number of practically plausible distributions. The baseline value for the co-primary outcome of cognition follows a $N(25, 9^2)$. The age of the patient at baseline is assumed to follow $N(70, 9^2)$, while the distribution for severity of the disease at baseline (mild=1/moderate=0) is Bernoulli(0.5). Finally, concomitant ACHEI or meantime use at baseline (Yes=1/No=0) follows a Bernoulli(0.75).

The mean and variance for age are chosen in line with knowledge about Alzheimer's disease. Once the covariates are generated, both responses: efficacy (change from the baseline for one of the co-primary endpoints) and tolerance are jointly generated from a multivariate normal distribution. The underlying models for both response variables are given below. For efficacy, this is

$$\begin{aligned}
Y_{ijk} &= \beta_1 B_{ij} + \beta_2 A_{ij} + \beta_3 M_{ij} + \beta_4 C_{ij} \\
&\quad +[\beta_5 I(k=1) + \beta_6 I(k=2) + \beta_7 I(k=3) + \beta_8 I(k=4) + \beta_9 I(k=5) + \beta_{10} I(k=6)] \\
&\quad \times I(T_{ij} = 140\text{mg}) \\
&\quad +[\beta_{11} I(k=1) + \beta_{12} I(k=2) + \beta_{13} I(k=3) + \beta_{14} I(k=4) + \beta_{15} I(k=5) + \beta_{16} I(k=6)] \\
&\quad \times I(T_{ij} = 100\text{mg}) \\
&\quad +[\beta_{17} I(k=1) + \beta_{18} I(k=2) + \beta_{19} I(k=3) + \beta_{20} I(k=4) + \beta_{21} I(k=5) + \beta_{22} I(k=6)] \\
&\quad \times I(T_{ij} = \text{Placebo}) \\
&\quad +b_i^Y + \varepsilon_{ijk}^Y. \qquad (1)
\end{aligned}$$

Here, Y_{ijk} is the efficacy response for patient j at site $i = 1, \ldots, 176$ and at visit $k = 1, \ldots, 6$. B_{ij}, A_{ij}, C_{ij}, M_{ij}, and T_{ij} are baseline, age, concomitant medication, MMSE, and treatment for patient j in site i, respectively. Further, $b_i^Y \sim N(0, \sigma^2)$ is a site-specific random effect, $\boldsymbol{\varepsilon}_{ij}^Y \sim N(0, \Sigma_{11})$ **is a random error vector for the efficacy response, the patient specific random effect can be absorbed into it by choosing** Σ_{11} **an unstructured** 6×6 **covariance matrix.** Similarly, the model for the tolerance response variable is:

$$\begin{aligned}
Z_{ijk} &= [\alpha_1 I(k=1) + \alpha_2 I(k=2) + \alpha_3 I(k=3) + \alpha_4 I(k=4) + \alpha_5 I(k=5) + \alpha_6 I(k=6)] \\
&\quad \times I(T_{ij} = 140\text{mg}) \\
&\quad +[\alpha_7 I(k=1) + \alpha_8 I(k=2) + \alpha_9 I(k=3) + \alpha_{10} I(k=4) + \alpha_{11} I(k=5) + \alpha_{12} I(k=6)] \\
&\quad \times I(T_{ij} = 100\text{mg}) \\
&\quad +[\alpha_{13} I(k=1) + \alpha_{14} I(k=2) + \alpha_{15} I(k=3) + \alpha_{16} I(k=4) + \alpha_{17} I(k=5) + \alpha_{18} I(k=6)] \\
&\quad \times I(T_{ij} = \text{Placebo}) \\
&\quad +b_i^Z + \varepsilon_{ijk}^Z. \qquad (2)
\end{aligned}$$

Now, Z_{ijk} denotes for the tolerance response for j^{th} patient at the ith site and at visit k, $\boldsymbol{\varepsilon}_{ij}^Z \sim$

$N(0, \Sigma_{22})$, ε_{ij}^{Z} **also is a random error vector for the tolerance response and here as well the patient-specific random effect is absorbed into it;** Σ_{22} is an unstructured 6×6 covariance matrix. The tolerance response variable is assumed to be related to treatment and visit. The joint distribution for both error terms is:

$$\begin{pmatrix} \varepsilon_{ij}^{Y} \\ \varepsilon_{ij}^{Z} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^{T} & \Sigma_{22} \end{pmatrix}\right],$$

with additionally Σ_{12} the 6×6 covariance structure between efficacy and tolerance. Two type of covariance structures were considered: (1) no correlation and (2) a correlation of 0.2 between efficacy and tolerance. For the latter, the covariance is assumed to have an autoregressive covariance structure in which the correlation between the measurements of efficacy and tolerance at the same visit is 0.2. The covariance matrices Σ_{11}, Σ_{22}, and Σ_{12} are given in the Appendix.

As far as our interest is in estimating the type I error rate, the data were generated under the null hypothesis of no difference in mean change between treatment arms at week 76 (visit 6), that is, $\beta_{10} = \beta_{16} = \beta_{22}$. The sets of parameters in Table 1 were used, and are chosen such that the measurements are within their appropriate ranges. A total of $S = 10,000$ datasets were simulated. Different set of parameters (treatment means) are considered, to study the sensitivity of the error rate to the choices of the parameter values, and to make sure the results are robust to different settings. Four sets of parameter values (treatment means) are considered, leading to four sets of simulations. The second set of parameter values are set with more rapidly declining rates whereas the third set of parameter values are set with more slowly declining rates in efficacy over time (at visit 6), when compared to the first parameter values. In the fourth setting, parameter values are chosen in such a way that the efficacy of the patients is declining very slowly over time and the difference among the three treatment groups is very narrow.

In addition, one simulation (Simulation 5) is considered under different switching criteria. Although the main interest is in assessing the type I error rate associated with the switching criteria explained in Section 3, it is useful to consider other switching criteria for comparison purposes, where the switching condition is highly related to the test statistics for the significance of the treatment effect. In this respect, the following criterion was used to switch to compare low dose group and placebo: whether the efficacy of the patients in the low dose group is better than that of patients in the high

dose group.

Table 1, About Here.

4.2 Incorporating Incompleteness

On a generated complete dataset, a missingness mechanism was applied. First, this was done under a MAR mechanism, where missingness depends on the observed values only:

$$\begin{aligned}&\text{logit}[P(D_{ij}=k|D_{ij}>k-1, y_{ij,k-1}, T_{ij}, Z_{ij,k-1})]\\ &= \psi_0+\psi_1 y_{ij,k-1}+\psi_2 I(T_{ij}=140\text{mg})+\psi_3 I(T_{ij}=100\text{mg})+\psi_4 Z_{ij,k-1}.\end{aligned}$$

Here, D_{ij} represents the time of drop out for patient j at site i, $y_{ij,k-1}$ is the previous observed measurement of this patient, representing the dependence of dropout on efficacy. Including treatment effect into the model helps ensure that the switching condition is satisfied by assigning a different probability of dropping out for different treatment groups. $Z_{ij,k-1}$ is the previous/observed tolerance response, representing the dependence of dropout on tolerance.

In the first batch of simulations, three sets of parameter values (Scenarios 1, 2, and 3), leading to a probability of switching of about 10%, 25%, and 50%, respectively, are considered. Scenario 1 introduces dropouts at a rate of about 47% in the high dose (140 mg) group, 39% in the low dose (100 mg), and 24% in the placebo group. It leads to a probability of switching of about 10%. Similarly, Scenarios 2 and 3 introduce dropout profiles of about (48.5%, 40.5%, 25%) and (50%, 43.5%, 26%) in the high dose, low dose, and placebo groups, respectively. Whereas the first three scenarios result in switching with a given non-zero probability, two other set of parameters (Scenarios 4 and 5), not resulting in switching, are also considered, for the sake of comparison. Scenario 4 introduces dropout of about 25% in the high dose group, 21% in the low dose group and 14% in the placebo group. In this case, dropout in the high dose group is not sufficiently large to meet the first requirement for switching. Scenario 5, on the other hand, introduces dropout of about 57% in the high dose group, 55% in the low dose group, and 48.5% in the placebo group. The percentage of dropout in all treatment groups is large but the percentage of dropout in the high dose group does not exceed that of placebo by 20% (second requirement for switching not satisfied).

To further see the impact of switching on the Type I error under a MNAR missing mechanism, such

a mechanism was also considered. This is done by including the current measurement, Y_{ijk}, into the logistic model:

$$\begin{aligned}&\text{logit}[P(D_{ij}=k|D_{ij}>k-1, y_{ij,k-1}, T_{ij}, Z_{ij,k-1})]\\ &\quad = \quad \psi_0+\psi_1 y_{ij,k-1}+\psi_2 y_{ijk}+\psi_3 I(T_{ij}=140\text{mg})+\psi_4 I(T_{ij}=100\text{mg})+\psi_5 Z_{ij,k-1}+\psi_5 Z_{ijk}.\end{aligned}$$

While it is evidently true that treatment is unknown to the trialist as well as to the patient, the patient does undergo the effects and therefore, in the true data generating model, a dependence on treatment is considered realistic, even though unavailable during the conduct of the trial, unless the trial is unblinded, or at least partially so (e.g., to the members of the monitoring committee).

Like the first five scenarios we considered under MAR, a corresponding collection of five scenarios under MNAR was also considered. Three sets of parameter values (Scenarios 6, 7, and 8), leading to probabilities of switching of about 10%, 25%, and 50% and two sets of parameters (Scenarios 9 and 10) that do not lead to switching. The percentage of dropout in each treatment arm for all scenarios is given in Table 2 and the ψ parameters are displayed in the Appendix. The final scenario corresponds to complete data, which is expected to produce a Type I error rate of 0.025. The latter is introduced as an internal checking device. These 11 scenarios are applied to the datasets generated under zero correlation between efficacy and tolerance. The same sets of parameters are also applied to data with a correlation of 0.2 between efficacy and tolerance, producing an additional 11 scenarios (Scenarios 12–22). An overview of the scenarios in the first batch of simulation is presented in Table 2.

Table 2, About Here.

The first set of simulations, which results from the first set of parameter values, is large, while the other four simulation batches, three with different sets of efficacy parameters (treatment means) and one with different switching criteria, are done on a smaller scale, i.e., only a subset of the scenarios from the first simulation is considered. The second batch of simulations encompasses the first 11 scenarios. The third batch contains 6 scenarios (1–5 and 11). The fourth batch contains only 4 scenarios; two scenarios with switching and the remaining two without switching. The fifth batch, with alternate switching criterion, encompasses two scenario with switching (1 and 3) and another

two without (4 and 5). To assess the sensitivity of the results to the choice of the number of sites as well as to the number of patients per site, two additional sets of simulations were conducted. These simulations are the same as the second batch of simulations, but with different numbers of sites and patients per site. In the first set of simulations, the number of sites is varied to 50, 176, and 300, thereby keeping the number of patients per site at 3. In the second batch, the number of patients per site is varied to 3, 10, and 30, keeping the number of sites fixed at 176.

Moreover, two additional simulations are conducted. These are similar to the second batch of simulations but the treatment effect is excluded when introducing missingness into the data. Therefore, the dropouts are entirely induced either by efficacy, tolerance, or both. Assume that efficacy can be positively or negatively related with dropouts. That is, patients may drop out due to poor efficacy at the previous visit ($\psi_1 > 0$) or due to temporary relief ($\psi_1 < 0$). We consider these cases, and some sub-cases, in turn. (1) The less efficacy there is, the more likely the patient drops out ($\psi_1 > 0$). In this case, the dropout has to be driven by tolerance. Otherwise, more missingness is likely in the placebo group and less in the high dose group and this implies that the switching condition cannot be satisfied. The tolerance response has no direct impact on the significance of the treatment effect but can have an indirect impact through the association with efficacy. (1.a) No correlation between efficacy and tolerance: Therefore, there is no association between significance of the treatment effect and switching criteria. As a result, no type I error rate inflation is expected. (1.b) Although not realistic, if we assume negative correlation between efficacy and tolerance, i.e., the higher tolerance response then the more efficacy (smaller value) there is, the switching criteria will have quite the opposite effect because we are switching to another treatment comparison when the chance of significance for the high dose is higher. Therefore, in that case, a lower type I error rate is expected. (1.c) Positive correlation between efficacy and tolerance, i.e., the higher tolerance response, the less efficacy (larger value) there is. This may result in inflation of the error rate. (2) The more efficacious the drug is, the more likely the patient drops out, stemming from temporary relief ($\psi_1 < 0$). In this case, if missingness is driven by efficacy, then the switching criteria does not systematically favor significance of the treatment effect. The switching criteria is associated with significance

of the treatment effect but in a quite opposite way. Also here, a lower type I error rate is anticipated. The two additional simulations correspond to these two situations, i.e., the first one when dropout is driven by efficacy and the second one when the dropout is induced by the tolerance response; a correlation between efficacy and tolerance of 0, 0.2, and 0.8 is considered.

4.3 Estimating the Type I Error Rate

The resulting datasets are analyzed using the aforementioned likelihood-based ignorable method, valid under MAR, thereby using all the available information without the need to either delete or impute measurements (Molenberghs and Verbeke 2005, Molenberghs and Kenward 2007). The same method was used to analyze the incomplete data resulting from MNAR. It enables us to see the impact of misspecification of the missing data mechanism. In addition to this likelihood-based ignorable method, last observation carried forward (LOCF) imputation was also used, limited to the case where site is considered a fixed effect. This allows us to assess the type I error rate under this more traditional analysis.

4.3.1 Scenarios with Switching

Those datasets satisfying the switching condition are analyzed based on the comparison of low dose and placebo (Model 1):

$$\begin{aligned} Y_{ijk} &= \beta_0 + \beta_1 I(T_{ij} = 100\text{mg}) + \beta_2 I(k=1) + \beta_3 I(k=2) + \beta_4 I(k=3) + \beta_5 I(k=4) \\ &\quad + \beta_6 I(k=5) + [\beta_7 I(k=1) + \beta_8 I(k=2) + \beta_9 I(k=3) + \beta_{10} I(k=4) \\ &\quad + \beta_{11} I(k=5)] I(T_{ij} = 100\text{mg}) + \beta_{12} A_{ij} + \beta_{13} M_{ij} + \beta_{14} C_{ij} + \beta_{15} B_{ij} + b_i + \varepsilon_{ijk}. \quad (3) \end{aligned}$$

The remainder are analyzed based on the comparison of high dose and placebo (Model 2):

$$\begin{aligned} Y_{ijk} &= \beta_0 + \beta_1 I(T_{ij} = 140\text{mg}) + \beta_2 I(k=1) + \beta_3 I(k=2) + \beta_4 I(k=3) + \beta_5 I(k=4) \\ &\quad + \beta_6 I(k=5) + [\beta_7 I(k=1) + \beta_8 I(k=2) + \beta_9 I(k=3) + \beta_{10} I(k=4) \\ &\quad + \beta_{11} I(k=5)] I(T_{ij} = 140\text{mg}) + \beta_{12} A_{ij} + \beta_{13} M_{ij} + \beta_{14} C_{ij} + \beta_{15} B_{ij} + b_i + \varepsilon_{ijk}. \quad (4) \end{aligned}$$

In both models, visit 6 is the reference time, hence β_1 is the parameter of interest. For each dataset, we either reject or do not reject the null hypothesis $H_0 : \beta_1 = 0$. The overall Type I error rate is

calculated as c/s, where c is the number of significant cases in both parts and s is the total number of simulations $(10,000)$.

4.3.2 Scenarios Without Switching

All datasets are analyzed based on the designed comparison of high dose and placebo. The type I error rate is calculated and compared with the type I error rate with switching. This enables us to see whether switching doses inflates the type I error rate and would need correction.

5 Simulation Results

Our primary goal was to compare the estimated Type I error rate with and without switching. A summary of the results of the first three batches, where site is considered as random and as fixed effect, is presented in Tables 3–5. The detailed results for all cases (site as fixed, random, and excluded) are presented in the Appendix. When site is excluded from the final analysis, the results produced an error rate much lower than the nominal level, but the type I error rates for the remaining three are quite similar. They approximately amount to the allowed type I error rate of 0.025, with minimum of 0.02 and maximum of 0.0305. The reason for the low Type I error rate for case c, site excluded from analysis, is because in generating the data, site has been considered, while in the final analysis it is neither included as fixed nor as random effect. Thus, as a consequence, a certain amount of variability is left unexplained. The immediate consequence is an overestimation of the standard errors, which in turn results in underestimating the Type I error rate. When site was excluded completely, both in data generating and final analysis, the expected type I error rate was obtained.

In general, there was not much difference between the scenarios with and without switching, in the estimated type I error rate **except for one scenario that will be discussed below**. This finding appeared when it was applied under the same missing data mechanism and correlation between efficacy and tolerance. There was no systematic trend in error rate when the probability of switching increases from 10% to 50%. It is also of interest to see how the type I error rate would vary if we assume different correlation between efficacy and tolerance. Noticeable differences in Type I error rate were not observed when a correlation of 0 and 0.2 between efficacy and tolerance was assumed

in generating the data. If we consider the datasets generated under MAR and MNAR and analyzed using direct likelihood, a technique valid only under MAR, then, although results were generally very similar between both, we noted that the type I error rate under MNAR is slightly higher in most of the scenarios when compared to its counterpart under MAR in the second batch of simulation.

Tables 3–5, About Here.

In contrast, and in line with expectation, the results obtained by analyzing these incomplete data using LOCF were very different. The type I error rate was huge under all scenarios. The reason is that the efficacy for the patients in the high dose group is better, i.e., smaller change/decline in the primary endpoint, than the efficacy of the patients in the control and low dose groups in the first five visits. Also, the efficacy in the low dose group is better than in the control group. Although the data are generated under the null hypothesis, implying that the efficacy of the patient is the same in all treatment arms at the last visit, carrying the observation forward for about 50% of the missing observations will inflate significance of the treatment effect at the last visit. Two factors determine how much the type I error rate inflates: (1) the magnitude of difference in the efficacy between the treatment arms and placebo; and (2) the difference in the percentage of missingness between the treatment arms and placebo.

Let us turn to the fourth batch. The error rate from the likelihood based analysis, like before, is controlled. The error rates from LOCF are still large. It is worth emphasizing that this large error resulted despite the data being generated under almost no treatment effect at any visit and very slight difference in the efficacy of the patients over time. This underscores just how much LOCF can inflate the error rate. Furthermore, the error rate is relatively lower when contrasted to their counterparts from the earlier results, with the same amount of missingness. This stems from the fact that the difference between the treatments arms and placebo is very small in this simulation. On the other hand, the difference in the percentage of missingness in the treatment arm forces the error rate to vary across the scenarios. For instance, the error rate under Scenario 4 is three times that of Scenario 3. Considering the difference in the percentage of missingness between the treatment arms and placebo explains why this huge difference can happen. Carrying observations forward assumes that patients had no further deterioration. Thus, more observations in the treatment group have

better efficacy than in the placebo group, leading to more spurious treatment effects.

Table 6, About Here.

The first four simulation results and the last two simulation (6 and 7) assess the error rate inflation associated with switching dose group based on the criteria explained in Section 3. Results indicated that the error rates are controlled under the likelihood method. Next, let us see the other switching criteria. Results presented in Table 7 clearly indicate the inflation of the type I error rate, which is in line with the theoretical expectation that the type I error rate can inflate when the switching criterion is correlated with the test statistic for significance. If the two treatment arms are compared with placebo independently, an error rate of approximately 0.025 is expected. When switching in a manner unrelated to the significance of treatment effect, we still have an error rate of 0.025 because of the random nature of the switching act. In the worst case scenario, however, as the switching criteria is related with the level of significance for treatment effect, situations with a significant treatment effect will be included, leading to an error rate of about 0.05. This is true, provided an appropriate method for handling missing data is used; combined with inappropriate methods, it can even inflate further, and considerably so, as shown using LOCF, for instance.

Table 7, About Here.

The sensitivity of the results to the choice of the number of sites, as well as to the number of patients per site was further assessed. The results, presented in Tables 8 and 9, show that it is insensitive to the number of sites and the number of patients per site. We do not see systematic differences in the Type I error rates between scenarios with and without switching, regardless the other settings. The bias of the parameters of two covariates, treatment difference at the last visit and MMSE score, for the different scenarios of one particular simulation is presented in Table 10.

Tables 8–10, About Here.

In the latter two simulations, the treatment effect is excluded when introducing missingness. Therefore, dropout is entirely induced either by efficacy or tolerance. In simulation 8 (Table 11), dropout is driven by efficacy such that the more efficacious the drug the

higher the dropout rate due to temporary relief. The type I error rate is approximately the nominal rate, no error rate inflation associated with switching is noticed. In simulation 9 (Table 12), dropout is driven by the tolerance response. More dropout is associated with a higher tolerance response. In this simulation, a correlation of 0, 0.2, and 0.8 between the efficacy and tolerance is considered. As before, the results show that, for both correlations 0 and 0.2, no type I error shift was noted. However, for the specific case of a very high correlation of 0.8, the switching had some impact. This is because the switching criterion is highly related with efficacy through the tolerance response.

Tables 11, and 12, About Here.

6 Concluding Remarks

Adaptive studies that allow switching between dose groups are routinely used. In particular, a design that allow to choose a dose group to compare to placebo based on dropout rate. That is, the primary analysis may switch or shift to a different treatment contrast if the dropout rate is too high. It is a concern whether the type I error rate inflates with such design and requires a multiplicity adjustment. In this study, a simulation experiment was set up to assess the type I error rate inflation, inspired by an Alzheimer's disease trial associated with switching dose level. The type I error rate was estimated treating site as fixed effect, random effect, as well as by excluding it from analysis. All of this was done under different correlation levels between efficacy and tolerance, and under different missing mechanisms.

Based on the analysis using an ignorable likelihood method, the estimated type I error rate with and without switching was approximately the nominal error rate for the different scenarios **except when dropout is strongly associated with efficacy** and it was insensitive to the choice of parameters. Using LOCF imputation, the error rate was inflated, both with and without switching. However, no type I error rate inflation associated with switching was observed. Under a switching criterion, related with the test statistic for treatment effect, type I error rate inflation associated with switching is noticed.

We conclude that, although switching doses in a data-driven fashion at the final analysis,

where the switching criteria is highly related with the primary endpoint, may in general lead to type I error rate inflation, the type I error rate inflation associated with switching was controlled for most scenarios for the Alzheimer trial with longitudinal outcome where patients are expected to worsen over time. An exception occurred for the specific case where dropout is strongly associated with efficacy. Therefore the switching criteria used need to be carefully studied regarding the possible association with significance of the treatment effect (Mallinckrodt et al 2003).

Our findings, when carefully consulted, can help minimize the impact of using untoward switch criteria.

Acknowledgment

Aklilu Ghebretinsae and Geert Molenberghs gratefully acknowledge support from IAP research Network P7/06 of the Belgian Government (Belgian Science Policy). T

References

Chang, M. (2008) *Adaptive Design Theory and Implementation using SAS and R*. Boca Raton: Chapman & Hall/CRC.

Chow, S.C. and Chang, M. (2008) Adaptive design methods in clinical trials—a review. *Orphanet Journal of Rare Diseases*, **3**, paper 11.

Mallinckrodt, C.H., Sanger, T.M., Dub, S., DeBrota, D.J., Molenberghs, G., Carroll, R.J., Potter, W.Z., and Tollefson, G.D. (2003) Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. *Biological Psychiatry*, **53**, 754–760.

Molenberghs, G. and Kenward, M.G. (2007) *Missing Data in Clinical Studies*. Chichester: John Wiley & Sons.

Molenberghs, G. and Verbeke, G. (2005) *Models for Discrete Longitudinal Data*. New York: Springer.

Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data.* New York: Springer.

Wang, M., Wu, Y.C., and Tsai, G.F. (2008) A regulatory view of adaptive trial design. *Journal of the Formos Medical Association*, **107**, 3–8.

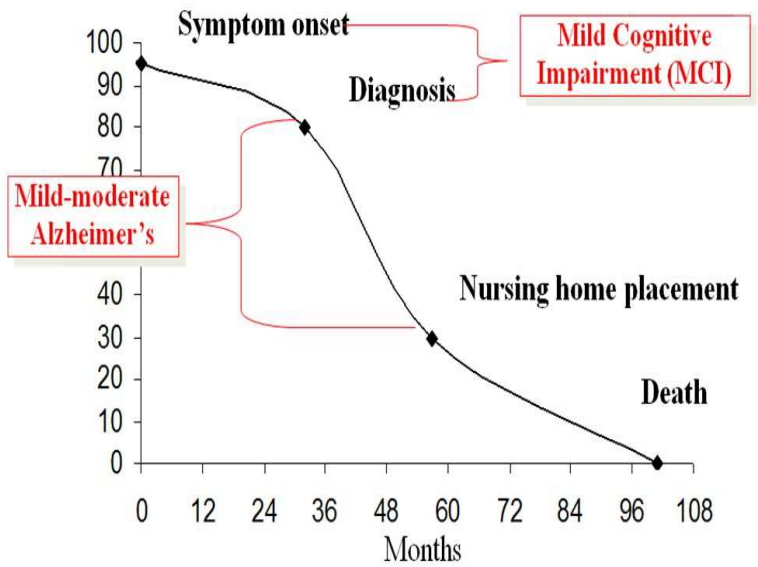

Figure 1: *Onset and expected decline of Alzheimer's disease/*

Table 1: *Simulation study. Parameter values used for data generation.*

Sim.	Efficacy	tolerance
1,5	$\beta_1 = 0.4$, $\beta_2 = 0.1$, $\beta_3 = 4$, $\beta_4 = -2$,$\sigma_y = 3$	$\sigma_z = 2$
	140 mg: $(\beta_5 - \beta_{10})$= 12, 16, 17,22, 25, 34	140 mg $(\alpha_1 - \alpha_6)$= 10, 12.2, 13.1, 14.5, 15, 15.5
	100 mg: $(\beta_{11} - \beta_{16})$= 15, 18, 20, 24, 27, 34	100 mg $(\alpha_7 - \alpha_{12})$= 8, 10.2, 11.5, 12.2, 13, 13.4
	Placebo: $(\beta_{17} - \beta_{22})$= 20, 24, 27, 31, 34, 34	Placebo $(\alpha_{13} - \alpha_{18})$= 3.5, 4, 4.2, 4.5, 4.8, 4.9
2,6,7,8,9	$\beta_1 = 0.4$, $\beta_2 = 0.1$, $\beta_3 = 4$, $\beta_4 = -2$,$\sigma_y = 3$	$\sigma_z = 2$
	140 mg: $(\beta_5 - \beta_{10})$ = 12, 14, 18, 24, 30, 36	140 mg $(\alpha_1 - \alpha_6)$= 11, 12, 13.2, 14.5, 15, 16
	100 mg: $(\beta_{11} - \beta_{16})$= 15, 16, 20, 26, 31, 36	100 mg $(\alpha_7 - \alpha_{12})$= 8, 9.2, 10.5, 12.2, 13, 13.4
	Placebo: $(\beta_{17} - \beta_{22})$= 20, 21, 25, 31, 36, 36	Placebo $(\alpha_{13} - \alpha_{18})$= 4, 4.5, 5.2, 6.5, 6.8, 7.2
3	$\beta_1 = 0.4$, $\beta_2 = 0.1$, $\beta_3 = 4$, $\beta_4 = -2$,$\sigma_y = 3$	$\sigma_z = 2$
	140 mg: $(\beta_5 - \beta_{10})$ = 12, 14, 17, 23, 27, 32	140 mg $(\alpha_1 - \alpha_6)$= 11, 12, 13.2, 14.5, 15, 16
	100 mg: $(\beta_{11} - \beta_{16})$= 15, 16, 19, 24, 28, 32	100 mg $(\alpha_7 - \alpha_{12})$= 8, 9.2, 10.5, 12.2, 13, 13.4
	Placebo: $(\beta_{17} - \beta_{22})$= 20, 21, 24, 28, 34, 32	Placebo $(\alpha_{13} - \alpha_{18})$= 4, 4.5, 5.2, 6.5, 6.8, 7.2
4	$\beta_1 = 0.4$, $\beta_2 = 0.1$, $\beta_3 = 4$, $\beta_4 = -2$,$\sigma_y = 3$	$\sigma_z = 2$
	140 mg:$(\beta_5 - \beta_{10})$=23, 23, 23.2, 23.4, 23.6, 24	140 mg $(\alpha_1 - \alpha_6)$=10, 12.2, 13.1, 14.5, 15, 15.5
	100 mg:$(\beta_{11} - \beta_{16})$=23, 23, 23.2, 23.5, 23.7, 24	100 mg $(\alpha_7 - \alpha_{12})$=8, 10.2, 11.5, 12.2, 13, 13.4
	Placebo:$(\beta_{17} - \beta_{22})$=23,23.2,23.2,23.6,23.7,24	Placebo $(\alpha_{13} - \alpha_{18})$=3.5, 4, 4.2, 4.5, 4.8, 4.9
8,9	$\beta_1 = 0.4$, $\beta_2 = 0.1$, $\beta_3 = 4$, $\beta_4 = -2$,$\sigma_y = 0.4$	$\sigma_z = 0.4$
	140 mg: $(\beta_5 - \beta_{10})$ = 12, 14, 18, 24, 30, 36	140 mg $(\alpha_1 - \alpha_6)$= 11, 12, 13.2, 14.5, 15, 16
	100 mg: $(\beta_{11} - \beta_{16})$= 15, 16, 20, 26, 31, 36	100 mg $(\alpha_7 - \alpha_{12})$= 8, 9.2, 10.5, 12.2, 13, 13.4
	Placebo: $(\beta_{17} - \beta_{22})$= 20, 21, 25, 31, 36, 36	Placebo $(\alpha_{13} - \alpha_{18})$= 4, 4.5, 5.2, 6.5, 6.8, 7.2

Table 2: *Overview of the scenarios considered for the first batch of simulations.*

					Av. % dropout		
Scenario	Corr.	Mech.	Switching	P(Switching)	high	low	placebo
1	0	MAR	Yes	0.10	47	39	24
2	0	MAR	Yes	0.25	48.5	40.5	25
3	0	MAR	Yes	0.50	50	43.5	26
4	0	MAR	No	-	25	21	14
5	0	MAR	No	-	57	55	48.5
6	0	MNAR	Yes	0.10	47	39	22.5
7	0	MNAR	Yes	0.25	48.5	40.5	23.5
8	0	MNAR	Yes	0.50	50	43.5	24.5
9	0	MNAR	No	-	25	21	12.5
10	0	MNAR	No	-	57	55	48.5
11	0	No	No	-	0	0	0
12	0.2	MAR	Yes	0.10	47	39	24
13	0.2	MAR	Yes	0.25	48.5	40.5	25
14	0.2	MAR	Yes	0.50	50	43.5	26
15	0.2	MAR	No	-	25	21	14
16	0.2	MAR	No	-	57	55	48.5
17	0.2	MNAR	Yes	0.10	47	39	22.5
18	0.2	MNAR	Yes	0.25	48.5	40.5	23.5
19	0.2	MNAR	Yes	0.50	50	43.5	24.5
20	0.2	MNAR	No	-	25	21	12.5
21	0.2	MNAR	No	-	57	55	48.5
22	0.2	No	No	-	0	0	0

Table 3: *Summary of the estimated type I error rate from the first batch of results, for site as a fixed and a random effect.*

				Estimated Type I error rate		
				Site fixed		Site random
Scenario	Corr.	Mech.	Switching	Ign. lik.	LOCF	Ign. lik.
1	0	MAR	Yes	0.024	1.000	0.025
2	0	MAR	Yes	0.023	1.000	0.025
3	0	MAR	Yes	0.027	1.000	0.027
4	0	MAR	No	0.026	1.000	0.022
5	0	MAR	No	0.026	1.000	0.029
6	0	MNAR	Yes	0.025	1.000	0.024
7	0	MNAR	Yes	0.023	1.000	0.025
8	0	MNAR	Yes	0.025	1.000	0.025
9	0	MNAR	No	0.023	1.000	0.022
10	0	MNAR	No	0.026	1.000	0.027
11	0	No	No	0.021	0.021	0.026
12	0.2	MAR	Yes	0.029	1.000	0.028
13	0.2	MAR	Yes	0.028	1.000	0.024
14	0.2	MAR	Yes	0.029	1.000	0.025
15	0.2	MAR	No	0.024	1.000	0.023
16	0.2	MAR	No	0.025	1.000	0.022
17	0.2	MNAR	Yes	0.025	1.000	0.028
18	0.2	MNAR	Yes	0.023	1.000	0.022
19	0.2	MNAR	Yes	0.024	1.000	0.022
20	0.2	MNAR	No	0.025	1.000	0.022
21	0.2	MNAR	No	0.025	1.000	0.025
22	0.2	No	No	0.024	0.024	0.027

Table 4: *Summary of the estimated type I error rate from the second batch of results, based on ignorable likelihood.*

				Estimated Type I error rate	
Scenario	Corr.	Mech.	Switching	Site fixed	Site random
1	0	MAR	Yes	0.0277	0.0264
2	0	MAR	Yes	0.0240	0.0251
3	0	MAR	Yes	0.0205	0.0249
4	0	MAR	No	0.0273	0.0250
5	0	MAR	No	0.0255	0.0250
6	0	MNAR	Yes	0.0274	0.0298
7	0	MNAR	Yes	0.0288	0.0251
8	0	MNAR	Yes	0.0220	0.0254
9	0	MNAR	No	0.0217	0.0267
10	0	MNAR	No	0.0271	0.0283
11	0	No	No	0.0234	0.0237

Table 5: *Summary of the estimated type I error rate from the third batch of results, based on ignorable likelihood.*

				Estimated type I error rate.	
Scenario	Corr.	Mech.	Switching	Site fixed	Site random
1	0	MAR	Yes	0.0248	0.0232
2	0	MAR	Yes	0.0204	0.0289
3	0	MAR	Yes	0.0288	0.0234
4	0	MAR	No	0.0295	0.0269
5	0	MAR	No	0.0259	0.0278
6	0	No	No	0.0260	0.0236

Table 6: *Summary of the type I error rate obtained from the fourth batch of simulations (with alternative parameter values); all scenarios are under MAR. (Av.per.drop.: average percentage of dropouts in the high, low and placebo response.*

			Estimated Type I error rate			
			Site fixed		Site random	
Scenario	Switching	Av.per.drop.	Likelihood	LOCF	Likelihood	LOCF
1	Yes	47,38,26	0.025	0.38	0.024	0.37
2	Yes	50,40,18	0.028	0.48	0.026	0.48
3	No	39.5,30,7	0.029	0.68	0.025	0.68
4	No	59,56,46.5	0.022	0.22	0.023	0.22

Table 7: *Summary of the type I error rate, obtained from the fifth batch of simulations with different switching criteria; all scenarios are under MAR and site is a fixed effect.*

Scenario	Switching	Type I error (likelihood)
1	Yes	0.03483
2	Yes	0.03320
3	No	0.02364
4	No	0.02899

Table 8: *Summary of the estimated type I error rate for the sixth batch of results based on ignorable likelihood for different number of sites, where site is a random effect.*

				Type I error for # sites		
Scenario	Corr.	Mech.	Switching	50	176	300
1	0	MAR	Yes	0.028	0.025	0.025
2	0	MAR	Yes	0.025	0.021	0.026
3	0	MAR	Yes	0.029	0.027	0.025
4	0	MAR	No	0.027	0.022	0.022
5	0	MAR	No	0.026	0.023	0.027
6	0	MNAR	Yes	0.023	0.022	0.030
7	0	MNAR	Yes	0.029	0.028	0.031
8	0	MNAR	Yes	0.030	0.029	0.030
9	0	MNAR	No	0.029	0.022	0.031
10	0	MNAR	No	0.026	0.023	0.029
11	0	No	No	0.028	0.020	0.030

Table 9: *Summary of the estimated type I error rate from the seventh batch of results, based on ignorable likelihood, for different number of patients per site, where site is a random effect.*

				Type I error for # pat.		
Scenario	Corr.	Mech.	Switching	3	10	30
1	0	MAR	Yes	0.025	0.029	0.027
2	0	MAR	Yes	0.021	0.029	0.026
3	0	MAR	Yes	0.027	0.029	0.024
4	0	MAR	No	0.022	0.029	0.030
5	0	MAR	No	0.023	0.026	0.023
6	0	MNAR	Yes	0.022	0.031	0.027
7	0	MNAR	Yes	0.028	0.029	0.025
8	0	MNAR	Yes	0.029	0.030	0.031
9	0	MNAR	No	0.022	0.029	0.026
10	0	MNAR	No	0.023	0.026	0.028
11	0	No	No	0.020	0.025	0.026

Table 10: *Summary of the bias and standard error results for the seventh batch of results based on ignorable likelihood.*

				bias (s.e.)	
Scenario	Corr.	Mech.	Switching	Treat. diff. at V6	MMSE
1	0	MAR	Yes	-0.00429(0.1370)	0.00336(0.1004)
2	0	MAR	Yes	-0.00341(0.1372)	0.00372(0.1005)
3	0	MAR	Yes	-0.00431(0.1369)	0.00277(0.1005)
4	0	MAR	No	-0.00232(0.1221)	0.00242(0.0962)
5	0	MAR	No	-0.00708(0.1595)	0.00323(0.1059)
6	0	MNAR	Yes	-0.00878(0.1362)	0.00287(0.1000)
7	0	MNAR	Yes	-0.00777(0.1360)	0.00388(0.1000)
8	0	MNAR	Yes	-0.00806(0.1359)	0.00185(0.1001)
9	0	MNAR	No	-0.00381(0.1215)	0.00225(0.0960)
10	0	MNAR	No	-0.01016(0.1597)	0.00343(0.1057)
11	0	No	No	-0.00034(0.1102)	0.00266(0.0920)

Table 11: *Summary of the estimated type I error rate from the Eighth batch of results, based on ignorable likelihood.*

Scenario	Corr.	Mech.	Switching	Type I error rate
1	0	MAR	Yes	0.022
2	0	MAR	Yes	0.026
3	0	MAR	Yes	0.025
4	0	MAR	No	0.025
5	0	MAR	No	0.025

Table 12: *Summary of the estimated type I error rate from the ninth batch of results, based on ignorable likelihood.*

				Type I error rate		
Scenario	Mech.	P(Switching)	Switching	corr=0	corr=0.2	corr=0.8
1	MAR	Yes	0.1	0.019	0.024	0.032
2	MAR	Yes	0.25	0.023	0.025	0.039
3	MAR	Yes	0.5	0.028	0.025	0.038
4	MAR	No		0.019	0.025	0.028
5	MAR	No		0.024	0.021	0.031

Appendix

A Covariance Structure Used for Efficacy and tolerance; Parameter Values Used for Introducing Missingness

$$\Sigma_{11} = \begin{pmatrix} 6 & 2.4 & 2.2 & 2.1 & 1.8 & 1.2 \\ 2.4 & 5 & 2.1 & 1.9 & 1.6 & 1.3 \\ 2.2 & 2.1 & 4.4 & 2.3 & 1.7 & 0.9 \\ 2.1 & 1.9 & 2.3 & 4.1 & 1.6 & 0.8 \\ 1.8 & 1.6 & 1.7 & 1.7 & 3.8 & 1.2 \\ 1.2 & 1.3 & 0.9 & 0.8 & 1.2 & 3.2 \end{pmatrix},$$

$$\Sigma_{22} = \begin{pmatrix} 1 & 0.4 & 0.32 & 0.31 & 0.28 & 0.21 \\ 0.4 & 1.2 & 0.6 & 0.42 & 0.38 & 0.24 \\ 0.32 & 0.6 & 1.3 & 0.5 & 0.44 & 0.32 \\ 0.31 & 0.42 & 0.5 & 1.1 & 0.56 & 0.38 \\ 0.28 & 0.38 & 0.44 & 0.56 & 1.2 & 0.42 \\ 0.21 & 0.24 & 0.32 & 0.38 & 0.42 & 1.4 \end{pmatrix},$$

$$\Sigma_{12} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\Sigma_{12} = \begin{pmatrix} 0.49 & 0.107 & 0.022 & 0.0041 & 0.00086 & 0.00018 \\ 0.089 & 0.49 & 0.102 & 0.019 & 0.0039 & 0.00085 \\ 0.016 & 0.091 & 0.48 & 0.088 & 0.018 & 0.004 \\ 0.0032 & 0.018 & 0.092 & 0.42 & 0.089 & 0.019 \\ 0.00062 & 0.0034 & 0.018 & 0.082 & 0.43 & 0.092 \\ 0.00011 & 0.00068 & 0.0033 & 0.015 & 0.078 & 0.42 \end{pmatrix}$$

$$\Sigma_{12} = \begin{pmatrix} 1.9370640 & 1.6952802 & 1.4024076 & 1.0250879 & 0.8569165 & 0.7342662 \\ 1.3847788 & 1.9091082 & 1.5695555 & 1.1424399 & 0.9439085 & 0.8086843 \\ 1.0140888 & 1.4055154 & 1.8308832 & 1.3327083 & 1.1029738 & 0.9365158 \\ 0.7686524 & 1.0671151 & 1.3872131 & 1.6068414 & 1.3250188 & 1.1279343 \\ 0.5832244 & 0.801930 & 1.052397 & 1.221365 & 1.619066 & 1.3825108 \\ 0.400769 & 0.5572728 & 0.733952 & 0.8530009 & 1.1404708 & 1.5724763 \end{pmatrix}.$$

B Result For the First Batch of Simulations

Table A.1: *Parameter values used for introducing missingness for Simulations 1 and 5.*

Scenario	Parameter values
1	$\psi_0 = -2.1$, $\psi_1 = -0.02$, $\psi_2 = 0.46$, $\psi_3 = 0.3$, $\psi_4 = 0.025$
2	$\psi_0 = -2.05$, $\psi_1 = -0.02$, $\psi_2 = 0.45$, $\psi_3 = 0.3$, $\psi_4 = 0.025$
3	$\psi_0 = -2$, $\psi_1 = -0.02$, $\psi_2 = 0.45$, $\psi_3 = 0.35$, $\psi_4 = 0.025$.
4	$\psi_0 = -2.75$, $\psi_1 = -0.02$, $\psi_2 = 0.3$, $\psi_3 = 0.2$, $\psi_4 = 0.025$.
5	$\psi_0 = -1.6$, $\psi_1 = -0.01$, $\psi_2 = 0.02$, $\psi_3 = 0.01$, $\psi_4 = 0.015$.
6	$\psi_0 = -2.2$, $\psi_1 = -0.01$, $\psi_2 = -0.01$, $\psi_3 = 0.45$, $\psi_4 = 0.3$, $\psi_5 = 0.015$, $\psi_6 = 0.02$.
7	$\psi_0 = -2.15$, $\psi_1 = -0.01$, $\psi_2 = -0.01$, $\psi_3 = 0.45$, $\psi_4 = 0.3$, $\psi_5 = 0.015$, $\psi_6 = 0.02$.
8	$\psi_0 = -2.1$, $\psi_1 = -0.01$, $\psi_2 = -0.01$, $\psi_3 = 0.44$, $\psi_4 = 0.35$, $\psi_5 = 0.015$, $\psi_6 = 0.02$.
9	$\psi_0 = -2.85$, $\psi_1 = -0.01$, $\psi_2 = -0.01$, $\psi_3 = 0.3$, $\psi_4 = 0.2$, $\psi_5 = 0.015$, $\psi_6 = 0.02$.
10	$\psi_0 = -1.6$, $\psi_1 = -0.005$, $\psi_2 = -0.005$, $\psi_3 = 0.02$, $\psi_4 = 0.01$, $\psi_5 = 0.005$, $\psi_6 = 0.01$.
11	-

Table A.2: *Parameter values used for introducing missingness for Simulations 2, 3, 6 and 7.*

Scenario	Parameter values
1	$\psi_0 = -3.76$, $\psi_1 = 0.015$, $\psi_2 = 0.51$, $\psi_3 = 0.36$, $\psi_4 = 0.05$
2	$\psi_0 = -3.72$, $\psi_1 = 0.015$, $\psi_2 = 0.51$, $\psi_3 = 0.37$, $\psi_4 = 0.05$
3	$\psi_0 = -3.675$, $\psi_1 = 0.015$, $\psi_2 = 0.51$, $\psi_3 = 0.36$, $\psi_4 = 0.05$
4	$\psi_0 = -4.45$, $\psi_1 = 0.015$, $\psi_2 = 0.4$, $\psi_3 = 0.28$, $\psi_4 = 0.05$
5	$\psi_0 = -2.93$, $\psi_1 = 0.015$, $\psi_2 = 0.013$, $\psi_3 = 0.011$, $\psi_4 = 0.05$
6	$\psi_0 = -4.1$, $\psi_1 = 0.01$, $\psi_2 = 0.01$, $\psi_3 = 0.6$, $\psi_4 = 0.41$, $\psi_5 = 0.025$, $\psi_6 = 0.025$
7	$\psi_0 = -4.06$, $\psi_1 = 0.01$, $\psi_2 = 0.01$, $\psi_3 = 0.6$, $\psi_4 = 0.42$, $\psi_5 = 0.025$, $\psi_6 = 0.025$
8	$\psi_0 = -4.01$, $\psi_1 = 0.01$, $\psi_2 = 0.01$, $\psi_3 = 0.597$, $\psi_4 = 0.42$, $\psi_5 = 0.025$, $\psi_6 = 0.025$
9	$\psi_0 = -4.8$, $\psi_1 = 0.01$, $\psi_2 = 0.01$, $\psi_3 = 0.45$, $\psi_4 = 0.3$, $\psi_5 = 0.025$, $\psi_6 = 0.025$
10	$\psi_0 = -3.2$, $\psi_1 = 0.01$, $\psi_2 = 0.01$, $\psi_3 = 0.024$, $\psi_4 = 0.018$, $\psi_5 = 0.025$, $\psi_6 = 0.025$
11	-

Table A.3: *Parameter values used for introducing missingness for Simulation 4.*

Scenario	Parameter values
1	$\psi_0 = -1.8$, $\psi_1 = -0.03$, $\psi_2 = 0.33$, $\psi_3 = 0.15$, $\psi_4 = 0.05$
2	$\psi_0 = -2.2$, $\psi_1 = -0.03$, $\psi_2 = 0.82$, $\psi_3 = 0.6$, $\psi_4 = 0.05$
3	$\psi_0 = -3.2$, $\psi_1 = -0.03$, $\psi_2 = 1.5$, $\psi_3 = 1.25$, $\psi_4 = 0.05$
4	$\psi_0 = -1.34$, $\psi_1 = -0.02$, $\psi_2 = 0.12$, $\psi_3 = 0.09$, $\psi_4 = 0.025$

Table A.4: *Parameter values used for introducing missingness for Simulation 8;* $\psi_1 y_{ij,k-1} + \psi_2 Z_{ij,k-1}$

Scenario	Parameter values
1	$\psi_0 = 1.86$, $\psi_1 = -0.12$, $\psi_2 = 0.02$
2	$\psi_0 = 1.86$, $\psi_1 = -0.118$, $\psi_2 = 0.02$
3	$\psi_0 = 1.86$, $\psi_1 = -0.117$, $\psi_2 = 0.02$
4	$\psi_0 = 1.2$, $\psi_1 = -0.125$, $\psi_2 = 0.02$
5	$\psi_0 = -1.2$, $\psi_1 = -0.02$, $\psi_2 = 0.02$

Table A.5: *Parameter values used for introducing missingness for Simulation 9;* $\psi_1 y_{ij,k-1} + \psi_2 Z_{ij,k-1}$

Scenario	Parameter values
1	$\psi_0 = -4.47$, $\psi_1 = 0.015$, $\psi_2 = 0.14$
2	$\psi_0 = -4.5$, $\psi_1 = 0.015$, $\psi_2 = 0.145$
3	$\psi_0 = -4.5$, $\psi_1 = 0.015$, $\psi_2 = 0.148$
4	$\psi_0 = -4.5$, $\psi_1 = 0.015$, $\psi_2 = 0.08$
5	$\psi_0 = -2.7$, $\psi_1 = 0.015$, $\psi_2 = 0.04$

Table A.6: *Site as fixed effect; likelihood.*

					Satisf. cond.		Not satisf. cond.		Total		
Scenario	Corr.	Mech.	Switch	P(switch)	Analyzed	Reject	Analyzed	Reject	Analyzed	Reject	Type I
1	0	MAR	Yes	0.10	959	24	8999	215	9958	239	0.024
2	0	MAR	Yes	0.25	2516	60	7526	175	10042	235	0.023
3	0	MAR	Yes	0.50	5590	157	4641	120	10231	277	0.027
4	0	MAR	No	-					9671	249	0.026
5	0	MAR	No	-					11380	300	0.026
6	0	MNAR	Yes	0.10	906	23	8975	226	9881	249	0.025
7	0	MNAR	Yes	0.25	2997	71	6986	158	9983	229	0.023
8	0	MNAR	Yes	0.50	5396	146	4659	105	10055	251	0.025
9	0	MNAR	No	-					9616	223	0.024
10	0	MNAR	No	-					11354	297	0.026
11	0	No	No	-					10158	213	0.021
12	0.2	MAR	Yes	0.10	1006	20	9040	269	10046	289	0.029
13	0.2	MAR	Yes	0.25	2448	69	7553	211	10001	280	0.028
14	0.2	MAR	Yes	0.50	5532	149	4510	139	10042	287	0.029
15	0.2	MAR	No	-					9616	234	0.024
16	0.2	MAR	No	-					11416	287	0.025
17	0.2	MNAR	Yes	0.10	932	21	8947	223	9879	244	0.025
18	0.2	MNAR	Yes	0.25	2946	65	6864	162	9810	227	0.023
19	0.2	MNAR	Yes	0.50	5435	148	4631	93	10066	241	0.024
20	0.2	MNAR	No	-					9574	235	0.025
21	0.2	MNAR	No	-					11430	288	0.026
22	0.2	No	No	-					10185	245	0.024

Table A.7: *Site as fixed effect; LOCF.*

					Satisf. cond.		Not satisf. cond.		Total		
Scenario	Corr.	Mech.	Switch	P(switch)	Analyzed	Reject	Analyzed	Reject	Analyzed	Reject	Type I
1	0	MAR	Yes	0.10	923	923	8045	8045	8968	8968	1
2	0	MAR	Yes	0.25	2298	2298	6684	6684	8982	8982	1
3	0	MAR	Yes	0.50	4913	4913	4119	4119	9032	9032	1
4	0	MAR	No	-					9909	9909	1
5	0	MAR	No	-					8784	8784	1
6	0	MNAR	Yes	0.10	891	891	8058	8058	8949	8949	1
7	0	MNAR	Yes	0.25	2416	2416	6724	6724	9140	9140	1
8	0	MNAR	Yes	0.50	4894	4894	4096	4096	8990	8990	1
9	0	MNAR	No	-					9776	9776	1
10	0	MNAR	No	-					8731	8731	1
11	0	No	No	-					10158	213	0.021
12	0.2	MAR	Yes	0.10	939	939	7961	7961	8900	8900	1
13	0.2	MAR	Yes	0.25	2326	2326	6708	6708	9034	9034	1
14	0.2	MAR	Yes	0.50	4983	4983	4010	4010	8993	8993	1
15	0.2	MAR	No	-					9916	9916	1
16	0.2	MAR	No	-					8793	8793	1
17	0.2	MNAR	Yes	0.10	874	874	8222	8222	9096	9096	1
18	0.2	MNAR	Yes	0.25	2770	2770	6350	6350	9120	9120	1
19	0.2	MNAR	Yes	0.50	4910	4910	4119	4119	9029	9029	1
20	0.2	MNAR	No	-					10040	10040	1
21	0.2	MNAR	No	-					8951	8951	1
22	0.2	No	No	-					245	10185	0.024

Table A.8: *Site as random effect.*

					Satisf. cond.		Not satisf. cond.		Total		
Scenario	Corr.	Mech.	Switch	P(switch)	Analyzed	Reject	Analyzed	Reject	Analyzed	Reject	Type I
1	0	MAR	Yes	0.10	997	22	9003	229	10000	251	0.025
2	0	MAR	Yes	0.25	2551	56	7449	191	10000	247	0.025
3	0	MAR	Yes	0.50	5527	150	4473	122	10000	272	0.027
4	0	MAR	No	-					10000	220	0.022
5	0	MAR	No	-					10000	288	0.029
6	0	MNAR	Yes	0.10	954	21	9046	220	10000	241	0.024
7	0	MNAR	Yes	0.25	2993	75	7007	171	10000	246	0.025
8	0	MNAR	Yes	0.50	5402	146	4598	99	10000	245	0.025
9	0	MNAR	No	-					10000	222	0.022
10	0	MNAR	No	-					10000	269	0.027
11	0	No	No	-					10000	261	0.026
12	0.2	MAR	Yes	0.10	990	24	9010	256	10000	280	0.028
13	0.2	MAR	Yes	0.25	2550	65	7450	177	10000	242	0.024
14	0.2	MAR	Yes	0.50	5537	143	4463	105	10000	248	0.025
15	0.2	MAR	No	-					10000	231	0.023
16	0.2	MAR	No	-					10000	221	0.022
17	0.2	MNAR	Yes	0.10	929	32	9071	249	10000	281	0.028
18	0.2	MNAR	Yes	0.25	2988	57	7012	158	10000	215	0.022
19	0.2	MNAR	Yes	0.50	5410	132	4590	89	10000	221	0.022
20	0.2	MNAR	No	-					10000	219	0.022
21	0.2	MNAR	No	-					10000	247	0.025
22	0.2	No	No	-					10000	274	0.027

Table A.9: *Site as random effect.*

					Satisf. cond.		Not satisf. cond.		Total		
Scenario	Corr.	Mech.	Switch	P(switch)	Analyzed	Reject	Analyzed	Reject	Analyzed	Reject	Type I
1	0	MAR	Yes	0.10	997	22	9003	229	10000	251	0.025
2	0	MAR	Yes	0.25	2551	56	7449	191	10000	247	0.025
3	0	MAR	Yes	0.50	5527	150	4473	122	10000	272	0.027
4	0	MAR	No	-					10000	220	0.022
5	0	MAR	No	-					10000	288	0.029
6	0	MNAR	Yes	0.10	954	21	9046	220	10000	241	0.024
7	0	MNAR	Yes	0.25	2993	75	7007	171	10000	246	0.025
8	0	MNAR	Yes	0.50	5402	146	4598	99	10000	245	0.025
9	0	MNAR	No	-					10000	222	0.022
10	0	MNAR	No	-					10000	269	0.027
11	0	No	No	-					10000	261	0.026
12	0.2	MAR	Yes	0.10	990	24	9010	256	10000	280	0.028
13	0.2	MAR	Yes	0.25	2550	65	7450	177	10000	242	0.024
14	0.2	MAR	Yes	0.50	5537	143	4463	105	10000	248	0.025
15	0.2	MAR	No	-					10000	231	0.023
16	0.2	MAR	No	-					10000	221	0.022
17	0.2	MNAR	Yes	0.10	929	32	9071	249	10000	281	0.028
18	0.2	MNAR	Yes	0.25	2988	57	7012	158	10000	215	0.022
19	0.2	MNAR	Yes	0.50	5410	132	4590	89	10000	221	0.022
20	0.2	MNAR	No	-					10000	219	0.022
21	0.2	MNAR	No	-					10000	247	0.025
22	0.2	No	No	-					10000	274	0.027

Table A.10: *Site excluded from the final analysis.*

					Satisf. cond.		Not satisf. cond.		Total		
Scenario	Corr.	Mech.	Switch	P(switch)	Analyzed	Reject	Analyzed	Reject	Analyzed	Reject	Type I
1	0	MAR	Yes	0.10	987	2	9013	6	10000	8	0.0008
2	0	MAR	Yes	0.25	2580	1	7420	4	10000	5	0.0005
3	0	MAR	Yes	0.50	5553	2	4447	5	10000	7	0.0007
4	0	MAR	No	-					10000	5	0.0005
5	0	MAR	No	-					10000	18	0.0018
6	0	MNAR	Yes	0.10	956	0	9044	6	10000	6	0.0006
7	0	MNAR	Yes	0.25	3033	0	6967	4	10000	4	0.0004
8	0	MNAR	Yes	0.50	5446	1	4554	5	10000	6	0.0006
9	0	MNAR	No	-					10000	6	0.0006
10	0	MNAR	No	-					10000	19	0.0019
11	0	No	No	-					10000	2	0.0002
12	0.2	MAR	Yes	0.10	984	2	9016	5	10000	7	0.0007
13	0.2	MAR	Yes	0.25	2545	3	7455	7	10000	10	0.0010
14	0.2	MAR	Yes	0.50	5532	3	4468	5	10000	8	0.0008
15	0.2	MAR	No	-					10000	2	0.0002
16	0.2	MAR	No	-					10000	20	0.0020
17	0.2	MNAR	Yes	0.10	945	2	9055	5	10000	7	0.0007
18	0.2	MNAR	Yes	0.25	2985	2	7015	5	10000	7	0.0007
19	0.2	MNAR	Yes	0.50	5434	3	4566	3	10000	6	0.0006
20	0.2	MNAR	No	-					10000	1	0.0001
21	0.2	MNAR	No	-					10000	15	0.0015
22	0.2	No	No	-					10000	0	0

Table A.11: *Site excluded in generating the data and in the final analysis.*

					Satisf. cond.		Not satisf. cond.		Total		
Scenario	Corr.	Mech.	Switch	P(switch)	Analyzed	Reject	Analyzed	Reject	Analyzed	Reject	Type I
1	0	MAR	Yes	0.10	993	23	9007	236	10000	259	0.026
2	0	MAR	Yes	0.25	2563	64	7437	209	10000	273	0.027
3	0	MAR	Yes	0.50	5520	144	4480	121	10000	265	0.027
4	0	MAR	No	-					10000	246	0.025
5	0	MAR	No	-					10000	258	0.026
6	0	MNAR	Yes	0.10	952	22	9048	237	10000	259	0.026
7	0	MNAR	Yes	0.25	2991	69	7009	192	10000	261	0.026
8	0	MNAR	Yes	0.50	5397	138	4603	107	10000	245	0.025
9	0	MNAR	No	-					10000	237	0.024
10	0	MNAR	No	-					10000	254	0.025
11	0	No	No	-					10000	244	0.024
12	0.2	MAR	Yes	0.10	990	35	9010	228	10000	263	0.026
13	0.2	MAR	Yes	0.25	2556	82	7444	192	10000	274	0.027
14	0.2	MAR	Yes	0.50	5476	158	4524	119	10000	277	0.028
15	0.2	MAR	No	-					10000	239	0.024
16	0.2	MAR	No	-					10000	255	0.026
17	0.2	MNAR	Yes	0.10	948	29	9052	225	10000	254	0.025
18	0.2	MNAR	Yes	0.25	2981	96	7019	177	10000	273	0.027
19	0.2	MNAR	Yes	0.50	5389	142	4611	119	10000	261	0.026
20	0.2	MNAR	No	-					10000	232	0.023
21	0.2	MNAR	No	-					10000	253	0.025
22	0.2	No	No	-					10000	242	0.024