RESPONSE PROFILES FOR LONGITUDINAL CLINICAL TRIAL DATA WITH SUBJECT DROPOUT UNDER REGULATORY CONSIDERATIONS

Craig H. Mallinckrodt[1], W. Scott Clark[1], Raymond J. Carroll[2], and Geert Molenberghs[3]

1. Eli Lilly & Co., Lilly Corporate Center, Indianapolis, IN 46285.
2. Dept. of Statistics, Texas A&M University, College Station TX.
3. Center for Statistics, Limburgs Universitair Centrum, Universitaire Campus, B-3590, Diepenbeck, Belgium

Key words: missing data, longitudinal data, mixed-effects models

Abstract

Treatment effects are often evaluated by comparing change over time in outcome measures. However, valid analyses of longitudinal data can be problematic, particularly when subjects dropout prior to completing the trial for reasons related to the outcome. In choosing the primary analysis for confirmatory clinical trials, regulatory agencies have for decades favored the last observation carried forward (LOCF) approach for imputing missing values. Many advances in statistical methodology, and also in our ability to implement those methods, have been made in recent years. The characteristics of data from acute phase clinical trials can be exploited to develop an appropriate analysis for assessing response profiles in a regulatory setting. These data characteristics and regulatory considerations will be reviewed. Approaches for handling missing data are compared along with options for modeling treatment effects and correlations between repeated measurements. Theory and empirical evidence are utilized to support the proposal that a likelihood-based mixed-effects model repeated measures (MMRM) approach, based on the missing at random assumption, provides superior control of Type I and Type II error when compared with the traditional LOCF approach, which is based on the more restrictive missing completely at random assumption. It is further reasoned that in acute phase clinical trials, unstructured modeling of time trends and within-subject error correlations may be preferred.

Introduction

Treatment effects are often evaluated by comparing change over time in outcome measures. However, valid analyses of longitudinal data can be problematic, particularly when subjects dropout prior to completing the trial for reasons related to the outcome (1, 2). Numerous methods for analyzing data in the presence of subject dropout have been proposed, examined, and implemented (1-18). The available methods are so numerous that choosing a suitable method can be difficult, especially given the importance of the inferences drawn from clinical trial data.

The health, happiness, and survival of millions of patients depend upon the safety and efficacy of medicinal products. Considerable unmet need exists in many therapeutic areas as patients wait for safer therapies, more effective therapies, or any therapy at all. The potential benefit from optimal analytic methods that maintain control of Type I and Type II error, especially in the presence of nonrandom dropout, highlights the importance of modeling decisions.

When determining a suitable approach to modeling longitudinal data, it is important to realize that no single "best" method currently exists. This implies that an analysis must be individually tailored for a given situation. It is therefore crucial that the desired attributes of the analysis are clear, and that the characteristics of the missing and nonmissing data are understood. We address these issues in the context of the regulatory environments in which new medicinal products are developed. Consequently, the objectives of this paper are: 1) to examine the characteristics of missing and nonmissing data that influence modeling decisions; 2) to examine the desired attributes of confirmatory clinical trial analyses in light of regulatory considerations; and 3) to propose an appropriate primary analysis for assessing response profiles in acute phase clinical trials. Although we concentrate primarily upon more general considerations, our proposals are illustrated via specific application – namely acute phase clinical trials testing therapies for major depressive disorder.

Data Characteristics

Missingness Mechanisms - Characteristics of the Missing Data

In many areas of clinical research, the consequences of missing data arising from subject dropout can be profound (2, 3, 4, 5). The potential impact missing data can have is dependent upon the missingness process (i.e., mechanisms) leading to the nonresponse. Data are classified as missing completely at random (MCAR) if the missingness does not depend on (is explained by) either the observed or unobserved outcomes. Data are missing at random (MAR) if the missingness depends on (is explained by) the observed outcomes. Data are missing not at random (MNAR) if missingness depends on (is explained by) the unobserved outcomes (4, chapter 6).

In the case of likelihood-based estimation, given that the parameters defining the measurement process (observed outcomes) are independent of the parameters defining the missingness process (unobserved outcomes), the missingness is said to be ignorable provided it arises from an MCAR or MAR process. Under these same conditions, however, missingness arising from an MNAR process is said to be nonignorable (6 p 218).

Subject dropout rates and patterns (reasons for missingness) in clinical trials may be affected by many factors, including the disease, the study population, the efficacy of treatments, side effects, and length of the trial. Frequently, subjects dropout for reasons that are related to the outcome of interest, and thus the data are not MCAR (6, p 229, 12, 13, 19). The MAR assumption is typically more plausible than the MCAR assumption (4, 6 p 239, 12, 13, 19) as the observed data explain much of the missingness in many scenarios. This may be particularly true in well-controlled studies, such as clinical trials, where extensive efforts are made to observe all the outcomes and the factors that influence them (16). Therefore, even though the objectives of clinical trials are not explicitly stated as such, it is fair to state that, by their very design, clinical trials seek to minimize the amount of nonignorable missingness (MNAR data).

Characteristics of the Nonmissing Data

Specific characteristics of the data from clinical trials may be as varied as the trials themselves. Nevertheless, many trials share certain general attributes that may be exploited when choosing an appropriate method for the primary analysis. For example, trials are often designed to yield reliable estimates of central tendency, and are therefore likely to yield reliable estimates of dispersion. Acute phase clinical trials in depression, for example, typically have observations taken on the primary outcome measure at 3 to 8 postbaseline time points. These measurement times are fixed, with narrow intervals. Thus measurements may be taken perhaps once per week, with the week 1 observation mandated to take place between days 5 and 9, the week 2 observation between days 12 and 16, and so on.

These general attributes suggest that it may not be necessary to model the within-subject error correlation structure arising from repeated measurement in these longitudinal trials in a parsimonious manner. This feature is particularly beneficial when, as is the case in depression trials, *a priori* knowledge suggests that the functional form of the response profiles can be difficult to anticipate, and linear time trends may not adequately describe the response profiles. Nonlinear trends may arise from inherent characteristics of the particular disease state and drug under study and/or from features inherent within the trial. In such cases, parsimonious approaches to modeling time trends may lead to inaccurate results.

Whenever the design of an experiment utilizes multiple measurements on the same subjects, modeling the (co)variance between the repeated measurements should be considered. Three principal sources of (co)variance should be considered (17): (i) Inter-individual variability (i.e., heterogeneity between individual profiles) is frequently an important component of the within-subject correlation structure. This variability may be attributable to inherent, subject specific factors, while additional association may arise from unaccounted for fixed effects; (ii) Random variation may arise due to time course error (serial correlation), which exists when residuals close to each other

in time are more similar than residuals further apart; (iii) Measurement errors may contribute to random variability. In the longitudinal setting, it is particularly important to consider the potential for measurement error variation to increase or decrease over time.

The relative importance of the various sources of (co)variance can be useful in guiding modeling choices for specific circumstances. For example, in analyses of objective physical measures such as blood pressure or laboratory values, subject specific factors may have the greatest contribution to within-subject correlations. In these cases, a compound symmetric (or random effects or random coefficients) structure may be appropriate since residuals have equal (or similar) correlations regardless of degree of adjacency. Unaccounted for fixed effects may also give rise to a compound symmetric structure. In analyses of subjective ratings, such as the Hamilton Depression Rating Scale (20), time course errors that decay with increasing distance in time (such as in an autoregressive structure) may also be important.

While it may well be possible in many situations to model the subject specific effects and serial correlation separately, this would be necessary only if interest existed in the subject specific effects. In clinical trials, focus is directed primarily toward the fixed effects. Therefore it may be equally appropriate, and more straightforward, to omit explicit modeling of the subject specific effects and model them as part of the within-subject errors. In such a case, the subject specific effects and serial correlation combine, with or without changes in measurement error variation over time, to yield an unstructured correlation pattern for the within-subject errors.

Analytic Approaches for Mising Data
Traditional Methods – MCAR Analyses

A common choice for the primary analysis in a variety of therapeutic areas is to assess mean change from baseline to endpoint via analysis of (co)variance with missing data imputed by carrying the last observation forward (LOCF). The LOCF approach assumes that missing data are MCAR and that subjects' responses would have been constant from the last observed value to the endpoint of the trial. These conditions seldom hold (6). Carrying observations forward may therefore confound treatment with time (2), leading to bias in estimates of differences between treatment groups in mean change from baseline to endpoint (Δ) and the associated standard errors (SE_Δ) (2, 6, 9, 10, 11, 12, 13, 19).

Despite these shortcomings, LOCF has long been the method of choice for the primary analysis in clinical trials intended to support registration of new medicinal products. This situation has arisen due to the simplicity of the LOCF approach, its ease of implementation, and the belief that the potential bias from carrying observations forward leads to a "conservative" analysis.

However, mean change from baseline to endpoint is only a snapshot view of the response profile of a treatment. Gibbons (2) stated that endpoint analyses, although common,

are insufficient since response over time must also be assessed in order to completely understand the efficacy profile of a given treatment. By its very design, LOCF change to endpoint is incapable of assessing response profiles over time.

MAR Methods

We previously noted that in many settings the MAR assumption is more reasonable than the MCAR assumption. An MAR method is valid if data are MCAR or MAR, but MCAR methods are valid only if data are MCAR.

Likelihood-based mixed-effects models offer a general framework from which to develop longitudinal analyses under the MAR assumption (6, 7, 8, 19), and can also be extended to MNAR. Using the MAR (or MNAR) framework, likelihood-based mixed-effects analyses are more robust to potential bias from missing data than LOCF (2, 6, 12, 13) and other MCAR methods such as generalized estimating equations. In mixed-effects analyses, information from the observed data is used to provide information about the missing data, but missing data is not explicitly imputed. These analyses are therefore easy to implement because no additional data manipulation is required to accommodate the missing data, and the analyses can be implemented using software (e.g. the SAS Procedure Mixed) that has been widely available for a number of years. Likelihood-based mixed-effects model analyses are examined in detail by way of example in a later section.

MNAR Methods

Methods that attempt to account for MNAR missingness simultaneously model the measurement process (observed outcomes) and the missingness processes (unobserved outcomes). While the potential advantages of MNAR approaches are clear, they require assumptions that cannot be validated from the data at hand (6 p 216). This, in turn, argues that for any specific scenario a definitive MNAR analysis does not exist and the appropriate statistical framework for implementing MNAR methods is that of sensitivity analysis (6, 19).

Regulatory Considerations

In confirmatory clinical trials intended to support registration of new medicinal products, results from the primary analysis are the foundation upon which decisions are made. Regulatory agencies typically require the primary analysis to be prespecified. This is almost universally interpreted as a single analysis, yielding one result, with all aspects specified in the protocol (multiple primary analyses may be prespecified with appropriate adjustments for multiplicity). Only in rare circumstances could results from an analysis specified as secondary in the protocol, or an analysis not specified in the protocol, overturn the results from the primary analysis – no matter how compelling the secondary analysis may be. Furthermore, regulatory agencies typically prefer the primary analysis to be based on the intent-to-treat principle and on a simple analytic method with a simple model.

The Analytic Dilemma

The advantages of a single, simple, prespecified primary analysis are compelling, especially when such analyses are familiar from long-standing use. However, this decision-making framework is problematic to implement in the presence of missing data. The simple, traditional analyses used in clinical trials are typically based on the unrealistic assumption that data are MCAR. Although the MAR assumption is more plausible than the MCAR assumption, the possibility of MNAR data is difficult to rule out. Unfortunately, MNAR methods can be complex, and are best implemented in a sensitivity analysis framework – which is inconsistent with the need for a simple, single analysis.

We propose that likelihood-based mixed-effects models resolve the analytic dilemma to the greatest extent possible and are an appropriate choice for the primary analysis. Multiple imputation (MI) is another MAR approach (4, 9, 14), but MI is not as well suited to a regulatory environment as likelihood-based mixed-effects analyses. Additional data manipulation and modeling choices are needed to impute the missing values and to account for the uncertainty of imputation. In contrast, the mixed-effects analyses can be dictated entirely by the design of the study, with no additional steps being required to accommodate the missing data.

An Example

A retrospective evaluation of a phase II clinical trial of an antidepressant is used to illustrate how data characters can be used to guide modeling decisions. Goldstein et al (21) summarized the details and results of this study. Consider first an informal examination of the reasons for discontinuation summarized in Table 1. Although, the percentages of patients completing the protocol in each treatment group were not remarkably different, patients receiving the experimental drug tended to dropout more frequently due to adverse events, whereas placebo-treated patients tended to discontinue more frequently due to lack of efficacy. While it is not possible to ascertain whether the dropouts due to conflict/other reasons and protocol violations should be considered MCAR or MAR, they were probably not MNAR since the cause of dropout was observed. It is possible that MNAR data arise from patients who were lost to follow-up. However, only 5% (1.4% + 8.6%)/2 of the patients in the example study were lost to follow up. It is interesting to note that loss to follow up occurred in approximately the same relative ratio between treatments as lack of efficacy. If loss to follow up was simply another manifestation of lack of efficacy, the observed data would at least partially explain the reasons for missingness due to loss to follow up and the impact of MNAR data would be less than suggested by the 5% overall loss to follow up.

A simple, formal test for missingness was implemented using logistic regression to determine whether the probability of dropout was influenced by the observed outcomes. Rate of improvement in the primary efficacy outcome strongly

influenced probability of dropout (p < .0001), suggesting that the MCAR assumption was not valid. Molenberghs et al (19) applied the Diggle and Kenward (18) selection model to test for the existence of MNAR data in this study. Modeling the missingness did not significantly improve the fit to the data, and thus no evidence for MNAR data was found. Therefore, characteristics of the missing data suggested an MAR method was a viable approach for the primary analysis of these data.

Table 1. Reasons for Study Discontinuation from an Antidepressant Clinical Trial.

Reason for discontinuation	Drug N=70 %	Placebo N=70 %
Protocol Complete	60.0	64.3
Lack of Efficacy	2.8	14.3
Adverse Event	15.7	4.3
Lost to Follow-Up	1.4	8.6
Conflict/Other	12.9	4.3
Protocol Violation	7.1	4.3

Consideration of the nonmissing data in this study suggested that a categorical modeling of time would involve less risk than fitting time as linear or polynomic. For example, antidepressants are often said to have delayed onset of action, which could result in nonlinear time trends. Furthermore, this trial employed titration from a potentially suboptimal dose to one which was 3-fold greater. This feature could also give rise to nonlinear responses over time.

It is reasonable to anticipate that correlations between the repeated measurements in antidepressant trials arise from both inter-subject variability and serial correlation. Between subject differences are likely to be important since not only is depression a notoriously heterogeneous disease, but serial correlation often arises from patients' and raters' familiarity with responses from previous visits. Furthermore, variability in response to experimental drug and placebo or active comparator results in variances increasing over time. Given that no hypotheses regarding subject specific effects were to be tested, these effects did not need to be explicitly modeled and could be modeled via the within-subject errors.

Based on these data characteristics, a specific version of likelihood-based mixed-effects model analyses, referred to as MMRM (mixed model repeated measures) was implemented as the primary analysis in subsequent trials of this molecule. Results of the subsequent trial were summarized by Goldstein et al (22).

Post-hoc analyses of the subsequent trial again provided strong evidence for violation of the MCAR assumptions. Molenberghs et al (19) reported that modeling the missingness significantly improved the fit to the data, and

thus potential evidence for MNAR data was found. However, these, and other authors (6 chapter 18, 18), have cautioned that evidence for or against MNAR data in the selection model framework has to be interpreted cautiously as a variety of confounding factors can influence the results.

Perhaps the most important result from the MNAR modeling of both the example study and the subsequent study was that the estimates of the treatment effects were not markedly different when applying an MAR vs. an MNAR analysis (19). That is, the MAR results were robust to the MNAR data, if in fact MNAR data existed. The robustness of the specific MMRM analysis to MNAR data is examined in detail in a subsequent section.

The MMRM analysis can be implemented via PROC MIXED in SAS (7). Postbaseline values or changes from baseline are assessed as the dependent variable. Specific details of MMRM are as follows: Independent variables include the fixed, categorical effects of treatment, time, and treatment-by-time interaction, along with the continuous effects of baseline and baseline-by-time interaction. An unstructured (co)variance matrix is used to model the within-subject errors, at least in preliminary analyses. Algorithms for selecting the structure that provides the best fit can be defined in the protocol. Parameters are estimated using Restricted Maximum Likelihood with the Newton-Raphson algorithm. A treatment-by-time interaction contrast is constructed to estimate the difference between treatments in mean change from baseline to endpoint. Asymptotically exact standard error (SE) and 95% confidence intervals (CI) are obtained as described by Littell et al. (7 p 499). Denominator degrees of freedom can be estimated using such methods as Satterthwaite's approximation (7 p 48) or others.

In this formulation of the mixed-effects analysis, the within-subject error correlation structure is used to model each subject's response profile as a deviation from the group mean. For example, assume subject X had outcomes worse than average at all visits until visit 4, where he dropped out. Least square means for subject X's group at visits 5 and 6 are adjusted to reflect subject X's poor performance. The magnitude of the "adjustment" is determined by the magnitude of the within-subject correlations and by the magnitude of subject X's deviation from the group mean. The "uncertainty" in the adjustment is determined by the amount of data contributing to the group mean, the amount of data on subject X, and the magnitude of the within-subject correlations.

Robustness of Likelihood-Based Mixed Effect Models to MNAR Data

Mallinckrodt et al (12, 13) assessed the robustness of the MMRM implementation of likelihood based mixed-effects analyses to MNAR data by comparing results from MMRM to results from an LOCF ANOVA approach in simulated data. The first study (12) compared the two methods in simulated scenarios in which a true difference between treatments in mean change from baseline to endpoint existed ($\Delta \neq 0$). The second study (13) focused on

Type I error rates by simulating scenarios in which the difference between treatments in mean change from baseline to endpoint was zero ($\Delta=0$). In both studies, data comparisons were made both before introducing dropout (complete data) and also in the same data sets after eliminating data in order to introduce MNAR missingness.

In analyses of complete data from both studies, estimates of Δ and SE_Δ from MMRM and LOCF were identical in each of the 120,000 data sets. For example, in the second study, pooled across all 32 simulated scenarios, the Type I error rate from each method was 5.26%; the average estimate of Δ was -0.01, which was not significantly different from the true value of 0.00 (p=.7942). Type I error rates in the 32 scenarios ranged from 4.47% to 5.63%. However, important differences existed between the methods when analysing data with dropout.

In the study where treatment differences at endpoint existed ($\Delta \neq 0$), estimates of treatment group differences in mean change from baseline to endpoint from MMRM were closer to the true value than estimates from LOCF in every scenario simulated. Pooled across all scenarios, confidence interval coverage was 94.24% and 86.88% for MMRM and LOCF respectively. LOCF underestimated the superiority of the superior treatment in some scenarios and overestimated superiority in other scenarios. Although LOCF is generally considered a conservative method, other scenarios in which LOCF is likely to overestimate the true treatment effect have been noted (5, 15).

In the Type I error rate study ($\Delta=0$), pooled across all scenarios, the Type I error rates for MMRM and LOCF were 5.85% and 10.36%, respectively. Type I error rates in the 32 scenarios ranged from 5.03% to 7.17% for MMRM, and from 4.43% to 36.30% for LOCF. Again, the MMRM results were robust to the presence of MNAR data. Greater inflation of Type I error in LOCF resulted from greater bias in estimates of mean change from baseline to endpoint and unduly small standard errors that were a consequence of failing to account for the uncertainty of imputation.

Discussion

We have noted that no universally best approach to analysis of longitudinal data with dropout exists. Characteristics of the missing and nonmissing data can guide the analyst to an appropriate choice for a given situation. In the context of clinical trials intended to support the registration of a new medicinal product, regulatory considerations suggest the primary analytical method should be simple, yield a single result, and use an entirely prespecified model. However, the simple analyses traditionally used in many longitudinal clinical trials are based on the unrealistic assumption that data are MCAR. Although the MAR assumption is more plausible than the MCAR assumption, the possibility of MNAR data is difficult to rule out. But MNAR methods can be complex, and are best implemented in a sensitivity analysis framework – which is inconsistent with the need for a simple, single analysis.

The traditional approach to dealing with this analytic dilemma has been to use the LOCF approach, under the assumption that while potentially biased by non-MCAR data, the bias led to a "conservative" analysis. In this context, conservative is typically thought of as not overestimating the treatment effect. The simulation studies cited herein (12, 13) illustrated, and other authors have noted (5, 6, 15, 19), that conservative behavior of LOCF is not guaranteed.

It is interesting to consider that the tendency to underestimate the superiority of a superior treatment (conservative bias) necessarily results in the tendency to underestimate the inferiority of an inferior treatment. Thus, such a bias would be conservative in the context of superiority testing, but would be anti-conservative for non-inferiority testing.

Additionally, if a method yielded biased estimates of treatment effects when treatment differences existed, what would happen when that method was used in a situation when no differences existed? When the true treatment difference was zero, bias would necessarily lead to nonzero estimates of treatment differences, and potentially inflate Type I error. Similarly, consider Alzheimer's disease, where the therapeutic aim is to delay or slow deterioration of mental status, as compared to situations such as depression where the goal is to improve the condition. If a treatment is in truth no more effective than placebo, but patients on the experimental therapy drop out due to adverse events, carrying the last observations forward assumes that these patients had no further deterioration in condition. Thus, carrying observations forward would lead to the false conclusion that drug was more effective than placebo.

Therefore, whether or not the bias is conservative may depend upon the disease state, the type of test, and on the true difference between treatments. These limitations suggest that the analytic dilemma for missing data is not resolved by a conservative method because conservatism cannot be guaranteed.

Likelihood-based mixed-effects models are based on reasonable assumptions regarding missing data, are robust to violations of those assumptions, and provide an appropriate general analytic framework for assessing response profiles in longitudinal clinical trials.

The flexibility of mixed-effects models can also be exploited when considering the attributes of the nonmissing data. It may be beneficial to model the within-subject error correlation structure and time trends in an unstructured manner. Although parsimonious modeling of time trends and correlations can be more efficient and more appropriate in some circumstances, the more general unstructured formulations may provide an additional degree of robustness to unexpected and unpredictable outcomes.

Likelihood-based mixed-effects analyses are consistent with regulatory needs for a simple, prespecified analysis, based on the intent-to-treat principle. For example, all details of the MMRM analysis (a specific version of mixed-effects analyses) are dictated by the design of the study, can easily be specified in the protocol, and are straightforward to implement with standard software. Molenberghs et al (19)

discuss why likelihood-based MAR methods are consistent with the intent-to-treat principle, and in fact are an improvement over LOCF in this regard, via use of all available data on all subjects.

Changes in primary analytic methodology may have implications for regulatory agencies and for the companies seeking the marketing registrations. Although those implications are beyond the scope of this paper, they warrant careful consideration. Nevertheless, the developments in statistical theory and in our ability to readily implement the theory may be too compelling to overlook.

Conclusion

No universally best approach to analysis of longitudinal data exists. However, likelihood-based mixed-effects analyses provide a powerful analytic framework that is well suited to regulatory environments, is resilient to biases from missing data, and provides flexibility for handling the unique circumstances of specific scenarios.

References

1. Milliken, G.A.; Johnson, D.E. *The Analysis of Messy Data: Designed Experiments,* Vol. 1; Chapman & Hall: New York, 1993; 326.

2. Gibbons, R.D.; Hedeker, D.; Elkin, I.; Waternaux, C.; Kraemer, H.C.; Greenhouse, J.B.; Shea, T.; Imber, S.; Stotsky, S.M.; Watkins, J.T. Some Conceptual and Statistical Issues in Analysis of Longitudinal Psychiatric Data. Arch. Gen. Psych. **1993,** *50*, 739-750.

3. Laird, N.M. Missing Data in Longitudinal Studies. Stat. Med, **1988**, *7*, 305-315.

4. Little, R.; Rubin, D. *Statistical Analysis with Missing Data.* John Wiley and Sons: New York, 1987.

5. Lavori, P.W. Clinical Trials in Psychiatry: Should Protocol Deviation Censor Patient Data. Neuropsychopharmacol **1992**, *6*, 39-48.

6. Verbeke, G.; Molenberghs, G. *Linear Mixed Models for Longitudinal Data.* Springer: New York, 2000.

7. Littell, R.C.; Milliken, G.A.; Stroup, W.W.; Wolfinger, R.D. The SAS system for mixed models. SAS Institute Inc: Cary, NC, 1996; Chap. 1-10.

8. Cnaan, A.; Laird, N.M.; Slasor, P. Using the General Linear Mixed Model to Analyze Unbalanced Repeated Measures and Longitudinal Data. Stat. Med **1997**, *16*, 2349-2380.

9. Lavori, P.W.; Dawson, R.; Shera, D. A Multiple Imputation Strategy for Clinical Trials with Truncation of Patient Data. Stat. Med. **1995**, *14*, 1913-1925.

10. Siddiqui, O.; Ali, M.W. A Comparison of the Random-Effects Pattern Mixture Model with Last-Observation-Carried-Forward (LOCF) Analysis in Longitudinal Clinical Trials with Dropouts. J. Biopharm. Stat **1998**, *8*, 545-563.

11. Heyting, A.; Tolboom, J.T.B.M.; Essers, J.G.A. Statistical Handling of Dropouts in Longitudinal Clinical Trials. Stat. Med **1992**, *11*, 2043-2061.

12. Mallinckrodt, C.H.; Clark, W.S.; David, S.R. Accounting for Dropout Bias Using Mixed-Effects Models. J. BioPharm. Stat. **2001**, *11*, 9-21.

13. Mallinckrodt, C.H.; Clark, W.S.; David, S.R. Type I error rates from mixed effects model repeated measures compared with fixed effects ANOVA with missing values imputed via LOCF. Drug Information J. **2001**, *35*, 1215-1225.

14. Rubin, D.; Shenker, N. Multiple Imputations in Health Care Databases: An Overview and Some Applications. Stat. Med **1991**, *10*, 585-598.

15. Little, R.; Yau, L. Intent-to-Treat Analysis for Longitudinal Studies with Drop-Outs. Biometrics **1996**, *52*, 1324-1333.

16. Rubin, D.B.; Stern, H.S.; Vehovar, V. Handling "don' t know" survey responses: the case of the Slovenian plebiscite. J. Am. Stat. Assoc. **1995**, *90*, 822-828.

17. Diggle, P.J.; Liang, K–Y.; Zeger, S.L. *Analysis of Longitudinal Data*; Clarendon Press: Oxford, England, 1994.

18. Diggle, P.J.; Kenward, M.G. Informative dropout in longitudinal data analysis (with discussion). Appl. Stat. **1995**, *43*, 49-93.

19. Molenberghs, G.; Thijs, H.; Carroll, R.J.; Kenward, M.G. Analyzing incomplete longitudinal clinical trial data (with discussion). Biometrics **2002**, submitted.

20. Hamilton, M. A rating scale for depression. J Neurol Neurosurg Psychiatr. **1960**, *23*, 56-62.

21. Goldstein, D.J.; Mallinckrodt, C.; Lu, Y.; Demitrack, M.A. Duloxetine in the treatment of major depressive disorder: A double-blind clinical trial. J Clin. Psychiatry **2002**, *63*, 225-231.

22. Detke, M.J.; Lu, Y.; Goldstein, D.J.; Hayes, J.R.; Demitrack, M.A. Duloxetine, 60 mg once daily, for major depressive disorder: A randomized double-blind placebo-controlled trial. J Clin Psychiatry **2002**, *63*, 308-315.