Published in final edited form as:

J Psychiatr Res. 2007 December ; 41(12): 1032–1041.

Statistical Analysis of Longitudinal Psychiatric Data with Dropouts

Sati Mazumdar, Ph.D.^{1,2}, Gong Tang, Ph.D.¹, Patricia R. Houck, MSH², Mary Amanda Dew, Ph.D.^{1,2}, Amy E. Begley, MA², John Scott, MA¹, Benoit H. Mulsant, M.D., M.S.^{2,3}, and Charles F. Reynolds III, M.D.²

¹Graduate School of Public Health, University of Pittsburgh

²Advanced Center for Interventions and Services Research for Late Life Mood Disorders and John A. Hartford Center of Excellence in Geriatric Psychiatry, Department of Psychiatry, University of Pittsburgh School of Medicine

³Centre for Addiction and Mental Health and Department of Psychiatry, University of Toronto

Abstract

Longitudinal studies are used in psychiatric research to address outcome changes over time within and between individuals. However, because participants may drop out of a study prematurely, ignoring the nature of dropout often leads to biased inference, and in turn, wrongful conclusions. The purpose of the present paper is: (1) to review several dropout processes, corresponding inferential issues and recent methodological advances; (2) to evaluate the impact of assumptions regarding the dropout processes on inference by simulation studies and an illustrative example using psychiatric data; and (3) to provide a general strategy for practitioners to perform analyses of longitudinal data with dropouts, using software available commercially or in the public domain. The statistical methods used in this paper are maximum likelihood, multiple imputation and semi-parametric regression methods for inference, as well as Little's test and ISNI (Index of Sensitivity to Nonignorability) for assessing statistical dropout mechanisms. We show that accounting for the nature of the dropout process influences results and that sensitivity analysis is useful in assessing the robustness of parameter estimates and related uncertainties. We conclude that recording the causes of dropouts should be an integral part of any statistical analysis with longitudinal psychiatric data, and we recommend performing a sensitivity analysis when the exact nature of the dropout process cannot be discerned.

Keywords

nonrandom missing; inferential methods; sensitivity analysis

1. Introduction

Longitudinal studies are widely used in psychiatric research to measure treatment outcomes over time within individuals as well as changes across these individuals. Despite the collection of multiple, repeated measurements over time in most longitudinal studies, some analyses use

Corresponding author: Name: Sati Mazumdar, Address: Department of Biostatistics, Graduate School of Public Health, 130 DeSoto Street, University of Pittsburgh, Pittsburgh, PA 15261, Telephone: 412-624-3023, Fax: 412-624-2183, E-mail: maz1@pitt.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

only baseline and endpoint values comparing pre-post changes across different groups by analysis of variance (ANOVA) or analysis of covariance (ANCOVA) methods. In such analyses, data are not utilized efficiently since measurements between the baseline and the endpoint are ignored. When endpoint values are not available due to dropouts, analysis with only complete cases and a traditional method called “last observation carried forward” (LOCF) are commonly practiced (Mallinckrodt et al., 2003). LOCF analysis uses all subjects and imputes the missing values with the last observed value, a method that assumes that the outcomes would not have changed from the last observed value. This assumption of constant profile after dropout is hard to justify in many treatment trial settings.

Random regression models (Gibbons et al., 1993) provide a powerful tool for analyzing longitudinal data. In statistical analysis, these models are also referred to as mixed effects regression models, mixed effects models, or multilevel models. Over the last decade, these models have been used frequently in the psychiatric literature (Mazumdar et al., 2002; Hedeker, 2003; Houck et al., 2004). These models use all the data that are collected longitudinally and also account for within-subject correlation. However, the use of random regression models can lead to invalid conclusions when data are not missing at random (Little and Rubin, 2002; Rubin, 1976); that is, when the reason for a subject to drop out depends on the underlying value of the unobserved outcome even given recorded data from that subject. The assumption of missing at random may not be correct in psychiatric studies because participants may drop out of the study prematurely because of treatment failure, side effects, or unwillingness to comply with the treatment protocol. In psychiatric studies, “respondent burden” that may arise from many visits and assessments is frequently seen to be a reason for non-compliance with study procedures, leading to withdrawal from the study. Unfortunately, investigating the causes of dropout as an integral part of data analysis is not yet a well established practice.

Various features of, and appropriate statistical analyses for, longitudinal data with dropouts have been reviewed and discussed in the statistical literature (Little, 1995). A recent tutorial article for handling dropouts in longitudinal studies was designed to synthesize and illustrate the broad array of techniques used to address outcome-related dropout mechanisms, with an emphasis on regression based methods (Hogan et al., 2004). Another report discussed progress in analyzing repeated measures data (Gueorguieva and Krystal, 2004) and concluded that repeated-measures ANOVA still continues to be used widely despite both problematic interpretation and the availability of other methods. They recommended mixed effects models as the preferred choice for the analysis of such data due to both their greater flexibility in modeling time effects or in properly accounting for correlation between repeated measurements and their ability to handle missing data more appropriately. The authors noted, however, that mixed effects models can give biased results in the presence of missing values that are missing not at random (MNAR). They concluded that specific guidelines for the use of various methods for analyzing data from clinical trials and longitudinal follow-up studies in psychiatry are not yet available.

Molenberghs et al. provided a comprehensive summary of modeling frameworks and methods for analyzing incomplete longitudinal clinical trial data in circumstances where data may be MNAR (Molenberghs et al., 2004). Schafer and Graham argued that since these analyses are problematic, it is unwise to blindly shift to such methods, just as it can be dangerous to ignore non-random mechanisms altogether (Schafer and Graham, 2002). They suggested the use of sensitivity analysis, a procedure to ascertain how a given model output depends upon the input parameters (Saltelli et al., 2002). This is a powerful paradigm for assessing the robustness and reliability of a fitted model.

Recent papers on longitudinal data analysis do not adequately address the needs of practitioners involved in psychiatric research to understand different dropout processes, associated

inferential issues, and the limitations of software available commercially or in the public domain. Although the models and the issues to be discussed are relevant to clinical trial data well beyond psychiatry, the specific application to psychiatric trials is of particular importance. In psychiatric trials, there are a wide range of reasons for dropout, including side effects, feeling better or feeling worse, onset of other psychiatric disorders, or withdrawal of consent due to a variety of personal reasons, some of which may be attributable to the presence of the psychiatric disorder being treated. These reasons, if not recorded as part of the observed data in the study, can lead to MNAR data, a possibility that should be considered in any psychiatric clinical trial.

Thus, in this article we describe several dropout processes and the associated inferential issues in non-technical language, and we present a general conceptual and statistical analytic framework in which these procedures can be followed to provide sensible conclusions.

Although psychiatric studies may include continuous and/or discrete outcomes, we consider only analytic methods pertaining to continuous outcome measures in order to focus the paper. We note that some of the methods discussed here can be extended to discrete data.

We illustrate these procedures with data from a longitudinal psychiatric clinical trial. We use simulation studies to demonstrate what may happen in the inferential domain if the underlying assumptions pertaining to a chosen method are not satisfied. We conclude with practitioner guidelines for analyzing longitudinal data with dropouts.

2. Description of Longitudinal Data with Dropouts

In typical longitudinal psychiatric studies, an outcome variable such as the Hamilton Rating Scale for Depression (HRSD) (Hamilton, 1960), is recorded repeatedly over time, usually in order to document change with treatment. The longitudinal nature of the resulting data set provides rich information, such as variability of the outcome within an individual, variability of the outcome between individuals, and trend of the outcome over time, as well as effect of treatment on outcome. However, dropouts are inevitable in most studies because some subjects may relocate, become frustrated with the perceived ineffectiveness of the assigned treatment, or experience severe side effects. Hence, the sub-sample with an outcome recorded at any given time may not be a random sample of the complete study population. The observed data set can be regarded as the result of a selection process. Because of potential selection bias, different dropout processes have different implications in terms of appropriate statistical analyses.

Observed data/intermittent missing data/dropouts

In the statistical literature, observed covariates and values of the outcome are called observed data; the missing values of the outcome from individuals who drop out prematurely are called missing data (Little and Rubin, 2002). We make an important distinction between “intermittently missing” and “dropout”. An individual may miss a particular visit for a specific reason and return at the next scheduled visit, i.e., “intermittently missing” or “not monotone missing” data. Or an individual may miss a visit at some point and never return (“dropout” or “monotone missing” data). The present article deals with the analysis of longitudinal data with dropouts. We summarize below the major dropout processes.

Missing Completely At Random (MCAR)

The dropout mechanism is called missing completely at random (MCAR) (Little and Rubin, 2002) or completely at random dropout (CRD) (Little, 1995) if the missingness does not depend on outcomes but perhaps on some covariates. An example of dropout dependent on a covariate might be a situation in which older subjects drop out of a treatment study at higher rates than younger subjects due to lower availability of transportation. This definition of MCAR is

common in practice and it is broader, for the sake of simplicity, than what is discussed by Little (1995) where covariate-dependent dropout was singled out as a separate dropout mechanism.

Missing At Random (MAR)

If the dropout mechanism depends only on the observed data (i.e., it is not associated with final outcome measured by the unobserved missing data), it is called missing at random (MAR) or random dropout (RD) (Little, 1995; Little and Rubin, 2002). For example, a subject may drop out from a study after some recorded improvement from the illness, feeling that further treatment is not necessary. We note that the MCAR process is a special case of the MAR process.

Missing Not At Random (MNAR)

For a general non-random dropout mechanism, the probability of dropout at a given time point may depend on the unobserved current value of the outcome, even given the observed data. Terms like outcome-dependent dropout, informative dropout (ID), parametric informative dropout, non-ignorable missing, or missing not at random, are seen in the statistical literature dealing with longitudinal and survival data (Hogan et al., 2004; Ten Have et al., 1998). To avoid potential confusion with the survival analysis literature, we refer to the general non-random dropout mechanism as “missing not at random (MNAR) (Hogan et al., 2004). A special case of MNAR is when the missingness depends on some unobserved characteristics of individuals such as underlying general health status, which is often modeled as random effects in mixed models (Little, 1995; Diggle and Kenward, 1994; Wu and Carroll, 1988; Wu and Bailey, 1989). In the present paper, we use the MCAR-MAR-MNAR terminology. We note that the CRD-RD-ID terminology used by Diggle and Kenward (1994) still exists and is coherent with the terminology used here.

A descriptive diagram

Suppose in a longitudinal study, covariates X are fully observed and outcome Y is scheduled to be recorded at K occasions as $Y_1, Y_2, \dots, Y_K$. However, dropout may occur and we denote by R_k the indicator of dropout at occasion k : $R_k = 1$ if Y_k is observed and $R_k = 0$ if Y_k is missing, $k = 1, 2, \dots, K$. Let Z be auxiliary variables, beside X and Y_k s, that may be related to the dropout process. At occasion k , suppose $(Y_1, Y_2, \dots, Y_{k-1})$ are observed. The following diagram then presents the typical case for MCAR, MAR and MNAR, respectively.

In the above diagram, associated variables are connected by a straight line. This diagram is similar to the one given in Schafer and Graham for the description of a general missing data mechanism (Schafer and Graham, 2002).

3. Inferential Procedures

Inferential procedures for the analysis of longitudinal data with dropouts can be grouped under three broad classes: (1) likelihood-based procedures, (2) multiple imputation-based procedures, and (3) semi-parametric regression based procedures. We will briefly describe these procedures here; for details we refer the reader to the statistical literature (Little, 1995; Diggle et al., 2002; Molenberghs et al., 2004).

Maximum likelihood (ML) based procedures

In the ML-based procedures, the response model and the dropout process model are both specified according to prior information or past experience. Inference is based on maximizing the likelihood function, given the observed data and the dropout indicators. Under general conditions, the estimates have the same properties as ordinary ML estimators. When the dropout process is either MCAR or MAR and parameters of the model for the repeated measures are not involved in the dropout process, the specification of a dropout model is not necessary and the inference is based on the observed values as if there were no dropouts. In ML, this paradigm is referred to as ignorable. For practitioners, this means that data can be analyzed by standard packages (Allison, 2000; Horton and Lipsitz, 2001; Collins et al., 2001). However, these software packages do not in general supply correct standard errors when data are MAR, even though the parameter estimates are unbiased. For example, the standard errors for the fixed effects supplied by PROC MIXED of SAS are under-estimated when data are MAR but not MCAR because it assumes that the fixed effects are orthogonal to the variance components which is incorrect under MAR (Kenward and Molenberghs, 1998; SAS Institute, 2004).

If the dropout process is MNAR, its specification is necessary for the ML estimation (Diggle and Kenward, 1994; Little, 1995). The computation is typically complicated and numerical approximations are required to calculate the likelihood function. Then inference is drawn through conventional likelihood procedures. In particular, likelihood ratio tests can be used to assess the appropriateness of the dropout process through comparison of the full model with a reduced model corresponding to an MAR process (Ten Have et al., 1998).

With appropriate prior distributions for the parameters of the complete-data model and the dropout model, corresponding Bayesian methods can also be utilized. In general, they produce similar inferences as the ML method under large sample size.

Multiple imputations (MI) based procedures

The multiple imputation (MI) method provides an inferential approach for a wide variety of missing data problems (Schafer and Graham, 2002). In MI, each missing value is replaced by a list of m (> 1) simulated values, and substitution of these for the missing values produces m plausible alternative versions of the complete data. Each dataset is analyzed using an appropriate complete data method and the results are combined to obtain overall estimates and standard errors that reflect missing-data uncertainty as well as finite-sample variation. Computations can be carried out by SAS PROC MI and MIANALYZE (SAS Institute, 2000; Yuan, 2000).

Three imputation methods are available in the MI procedure. The choice of methods depends on the type of missing data pattern. For monotone missing data patterns (no intermittent missing data before a dropout), either a parametric regression method or a weighting method that uses propensity scores is appropriate. For an arbitrary missing data pattern, a Markov Chain Monte Carlo (MCMC) method (Schafer, 1997) can be used. Brief descriptions of these methods are given below.

In the parametric regression method, a model is fitted for each variable with missing values, using the other variables as covariates. The resulting model is used to impute the missing values (Little and Rubin, 2002). The propensity score weighting method requires the calculation of a propensity score, which is defined as the conditional probability of missingness or dropout given a vector of observed covariates (Rosenbaum and Rubin, 1984). At each occasion of dropout, propensity scores are generated for all observations that are subject to dropout at the occasion. Then, all observations are grouped based on these propensity scores. Finally, an

appropriate Bayesian bootstrap imputation is applied within each group (Lavori et al., 1995). The MCMC method generates random draws from intractable probability distributions via Markov chains (sequences of random variables in which the distribution of each value depends on the previous value) to create replicates of the distributions of interest from which inferences are made (Schafer, 1997). Hence, it should be noted that the MI based procedures can be framed both within frequentist and Bayesian frameworks. A drawback with MI is that the imputation procedure is often inconsistent with the assumed model for the repeated measure process and may cause bias.

Methods using generalized estimating equations and inverse probability weighted estimating equations

Sometimes the nature of the correlation among repeated measures is unclear or too complicated to specify. In such a situation, generalized estimating equations (GEE) provide a robust inferential procedure (Zeger and Liang, 1986). This method is especially useful when the response data are not normal but come from an exponential family. Missingness is ignorable in GEE only when data are MCAR. For other dropout processes (MAR and MNAR), a semiparametric method with inverse probability weighted estimating equations has been developed, where observed responses are weighted by the inverse of the estimated probability with which these values are observed (Horvitz and Thompson, 1952; Robins et al., 1995). This method allows great flexibility in the choice of the response model. However, the dropout process must be correctly specified to avoid biased estimates.

For a simple bivariate data set with a covariate and an outcome variable, we are usually interested in the prediction model of the outcome by the covariate. If the covariate is fully observed but the outcome is subject to missingness and the missing-data mechanism only depends on the outcome, the complete cases with both variables observed constitute a random sample for the conditional distribution of the covariate. Based on this observation, a pseudolikelihood method can be used to make inferences based on the prediction model, and one can avoid specifying a nonignorable missing-data mechanism (Tang et al., 2003; Tang et al., 2004). Application of this method to repeated-measures data remains under development and the details will not be discussed here.

4. Statistical Models for Data with Dropouts

The most commonly used statistical models for the analysis of longitudinal data with dropouts are selection models and pattern-mixture models. These models differ in the way the joint distribution of the outcome variable and the dropout indicator is factored. Selection models assume one model for the outcome process and another model for the dropout process with covariates, and use the outcomes as potential predictors of dropouts. According to specific features of the outcome variable and the dropout mechanism, selection models may use any of the inferential procedures described in section 3.

Pattern-mixture models stratify the observed data according to the dropout patterns and model the response process within each pattern by different models. “Dropout pattern” here refers to the number and timing of missing values. Data for each dropout pattern are modeled separately, and the parameter estimates for each dropout pattern are then combined using some weighting scheme (Hedeker and Gibbons, 1997). Pattern-mixture models are easier to implement than selection models, but they often result in a problematic interpretation of the outcome process. For example, the size of the treatment effect may differ for patients who belong to different dropout patterns. The overall treatment effect would be the weighted average of these effects over all dropout patterns. However, the differences in the corresponding effect sizes may not be negligible if studies using different methods are used in a meta-analysis. There are some

recent applications of pattern-mixture models with sophisticated analytical procedures (Shen and Weissfeld, 2005a; Shen and Weissfeld, 2005b).

For random-effects dependent dropouts, a special case of the MNAR process, flexible shared-parameter models have been described in which the random effects are regarded as shared random parameters between the response model and the dropout model (Little, 1995; Diggle and Kenward, 1994; Ten Have et al., 1998). The estimation and inference are often based on an approximated likelihood function since the integration involved in the likelihood function is often intractable.

For the general non-random dropout process, the process is usually not well understood; that is, the specific functional form for the dropout model is unknown. However, misspecification of the dropout model often leads to biased estimates of parameters from the response model. In practice, we recommend sensitivity analysis to investigate the impact of alternative assumptions about the dropout process on parameter estimates for the response model.

5. Detection of the Dropout Mechanism

In any given study, the nature of the dropout mechanism can be detected to a large extent by a review of the reasons for dropout. For psychiatric clinical trials, this procedure, if implemented in the protocol, will not pose much burden to the data collecting operation. A critical examination of the reasons for dropouts will provide insight into the conduct and follow-up procedures of the trial and will complement results from any sensitivity analysis.

Another approach to detection is Little's test, which assesses whether the missing data are MCAR or not (Little, 1988). The test divides the sample into groups based on the patterns of missingness for the outcome variable and is based on how pattern-specific means differ from the pooled estimates of the corresponding means under the MCAR assumption. A step-by-step procedure for this test is available, along with SAS coding (Fairclough, 2002).

A sensitivity analysis can also aid in the detection of the dropout mechanism. Verbeke and Molenberghs (2000) defined a sensitivity analysis as one in which several statistical models are considered simultaneously and/or where a statistical model is further scrutinized using specialized tools (such as diagnostic measures). The simplest procedure is to fit a selected number of (nonrandom) models that are all deemed plausible. The inference drawn from these models shall give investigators a good idea about the range of the parameter estimates and how likely it is that an effect of interest is real. The models can be based on various dependencies of the missing data process on the outcomes.

Sensitivity analysis can be based on local influence (Cook, 1986; Chatterjee and Hadi, 1988). Examples of how influence statistics can be used to assess the potential impact of nonrandom missingness are given in Verbeke and Molenbergh (2000). The general idea in influence analysis is to assess how observations, individually or collectively, excessively influence the fitted model as compared to other observations in the data set. In the present paper, we are interested in the influence the nonrandomness of dropout exerts on the parameters of interest, which will most often be the fixed effects parameters. This can be examined by considering a dropout model, e.g., a logit model, in which different subjects give different weights to the outcome at a particular occasion to predict dropout at that occasion. If all of these weights are equal to zero, the model becomes an MAR model. By perturbing these weights around zero in an influence analysis (Cook, 1986), the possibility of a nonrandom model is investigated.

The *Index of Sensitivity to Ignorability* (ISNI) is another tool that measures sensitivity to key inferences in the neighborhood of the ignorable missing at random (MAR) mechanism and provides a practical approach to assess the need for more complex nonignorable models at an

early stage of analysis (Troxel et al., 2004; Ma et al., 2005). The index measures the extent to which small departures from ignorability affect the ML estimate of the parameter of interest. Because ISNI depends on the units of measurement, a scale-free measure, called the “sensitivity transformation c ,” has been proposed (Troxel et al., 2004). Large values of c suggest that sensitivity occurs only in cases of extreme nonignorability, whereas small values suggest that sensitivity may be a problem even when the nonignorability is modest. Troxel et al (2004) have suggested using $c < 1$ as a cutoff value to identify important sensitivity.

6. A Useful Framework

Figure 1 presents a flowchart giving a general framework for approaching the analysis of longitudinal data with dropouts. If there is no strong clinical evidence that supports an MNAR dropout process, SAS PROC MIXED is recommended for estimation. However, its standard errors for fixed effects are underestimated for an MAR process. Under this circumstance, OSWALD may be used for valid inference or further programming is required. Nevertheless, it is also useful to conduct some sensitivity analysis such as exploring estimates under some plausible MNAR processes or ISNI analysis to get a broader picture regarding the impact of plausible MNAR processes.

7. Illustrative Example

Data from a clinical trial for the treatment of geriatric depression are used in this paper as a worked example (Mulsant et al., 2001) in order to illustrate empirically the different statistical models that are described in earlier sections. Results from this example also provide us with realistic parameter values for the simulation study. The sample comprised 116 geriatric inpatients and outpatients who met DSM-IV criteria (First et al., 1997) for a major depressive episode, without psychotic features or a past history of bipolar or schizoaffective disorder. An entry HRSD score of 15 or above and a Mini-Mental State Examination (Folstein et al., 1975) score of 15 and above were required for study entry. Subjects with contraindications to treatment with either of the protocol medications (nortriptyline and paroxetine) were excluded.

Qualified patients who consented to participate were randomly assigned to either nortriptyline (NT) ($n=54$) or paroxetine (PX) ($n=62$), and dosing was titrated according to protocol. Depression severity was assessed weekly for 12 weeks with the HRSD. Patients who failed to tolerate the randomized treatment or did not show satisfactory progress were terminated from the randomized portion of the study and were treated openly with the other protocol drug or other drugs as determined by the treating physician. The monitoring of depression severity was continued during the open treatment. It was suspected that tolerance to treatment is associated with depression severity of a patient.

A total of 58 of the 116 subjects (50%) completed the 12-week randomized treatment. About 30% of the subjects were terminated from the randomized protocol by week 6. Forty-five percent of the subjects treated with NT and 55% of those treated with PX completed the trial. Side effects led to dropout from the randomized treatment in 28 subjects (18 NT; 10 PX); 9 subjects withdrew consent (4 NT; 5 PX); 5 were noncompliant (1 NT; 4 PX) and the rest of the dropouts withdrew for other reasons, including new medical conditions and burden to the subject (e.g. inability to arrange for transportation). Figure 1 shows the mean HRSD score profiles by treatment groups. It includes number of subjects with measured outcome values for each week. Though the profiles seem to be similar for the two treatments a differential dropout pattern is seen between the two groups.

We analyzed this data set using a mixed model that included an intercept term, a treatment group effect, linear time effect, and a group by time interaction effect. The intercept and the time effects were considered as random. We assumed different dropout processes in our

analyses. As mentioned earlier, this strategy served two purposes. First, it allowed us to illustrate the use of the statistical models and different inferential procedures that were described earlier as a sensitivity analysis. Second, it helped us to design our simulation study. We summarize the results of our sensitivity analysis in Table 1. We note that the model for the outcome remained the same in all analyses.

The first column gives the parameters of the mixed model. The second column gives the parameter estimates for a selection model when an MAR process was assumed and PROC MIXED was used (SAS Institute, 2004). This procedure assumes the intermittent missing values as MCAR. We used the ML inferential procedure. The third column gives the results under the same assumptions (MAR) but using the multiple imputation inferential procedures for the parameter estimates. We first used the MCMC method for multiple imputation using PROC MI, modeled the data with PROC MIXED, and then used PROC MIANALYZE for summarizing the results.

The fourth column gives the results under an MNAR process using a logistic dropout process (Ten Have et al., 1998). We included this model mainly due to its convenience and feasibility. We assumed that the logit of the probability of dropout at a particular time is a linear function of the unobserved response at that time and the observed response at the previous time, and we used OSWALD for modeling (Smith et al., 1997). Intermittently missing values were regarded as MCAR. Because OSWALD does not provide standard error estimates, we employed a nonparametric bootstrap procedure to calculate the standard errors. Specifically, individuals were randomly sampled with replacement to form bootstrap samples with the same number of individuals as the original data and OSWALD was applied on each sample. Then the bootstrap standard errors were calculated based on the parameter estimates from 100 bootstrap samples (Begley et al., 2005).

The fifth column presents the random-effects pattern mixture model using SAS codes. We divided the subjects into groups based on their missing-data patterns that were then incorporated as main effects in our mixed model using dummy variables. We included the interaction terms of these dummy variables with the other variables in the original mixed model that provided sub-models for each of the missing data patterns. Finally, the overall averaged estimates for the model parameters were derived (Fairclough, 2002).

The sixth column provides the results of a simple version of the inverse probability weighted estimating equations (Robins et al., 1995). This was done by weighting the complete cases by the inverse of the probability of being a complete case. For each timepoint after baseline, a logistic regression was fit to predict the dropout probability using observed covariates and responses. Then for each complete case, the probability of being a complete case is estimated by the product of the predicted dropout probabilities at all timepoints. The code for this procedure was written in SAS by the authors.

The first three selection models (columns 2, 3, and 4) show comparable results with significant time effects only and similar standard errors. The observed differences between models are negligible given the magnitude of the standard errors. Results from the pattern mixture model (column 5) and the semi-parametric regression model (column 6) are also comparable to the selection model results. However, a smaller time effect with smaller standard error is noticed. Although the numerical values of the estimates vary, the inferences are the same with a statistically significant time effect. However, the differences in the corresponding effect sizes may not be negligible if studies using different methods are used in a meta analysis.

We have summarized the reasons for dropouts in Table 2. It is seen that the main reason for dropout was side effects (more so in the NT group than the PX group) while only a few subjects dropped out for psychiatric or medical reasons. The reason for dropout can be biased as it may

depend on the symptom level or the unobserved outcome value of the subject after dropping out. This is seen frequently in psychiatric trials and provides a bigger challenge for researchers to account for it in the modeling process.

Results from Little's test did not support an MCAR assumption over an assumption of MAR. For the ISNI analysis, we used a logistic dropout model (Troxel et al., 2004). The sensitivity transformation c is found to be greater than 1 for all the parameter estimates indicating that the missingness can be considered to be ignorable. In summary, we conclude that the MAR assumption is reasonable and provides valid parameter estimates in the present example.

8. Impact of Dropout Mechanism on Inference: Simulation Studies

To illustrate how different assumptions about the dropout mechanism affect the results of an analysis, we simulated a data set using a scenario that is common in psychiatric efficacy trials as seen in the illustrative example presented in Section 7. The simulated data set represented a two group longitudinal study over a 12-week period. Weekly observations were generated using a mixed model that included an intercept term, a group effect and a linear time effect and an MNAR process. We did not use a time by group interaction effect to keep the simulations simple. The MNAR process was implemented by a logistic dropout model which, in general, assumes that the logit of the conditional probability of dropping out at a particular time point is a linear function of outcome values at previous times, the unobserved outcome at the specific dropout time, and other potential covariates (Smith et al., 1997). To keep the simulations simple, we assumed that the logistic dropout model depends only on the observed outcome just before the dropout time and the unobserved outcome at the dropout time. Denoting by y_j , the outcome at time j ($j = 1, 2, \dots, k$) the dropout model can be written as,

$$\log it[p_k(y_1, y_2, \dots, y_{k-1}, y_k)] = \theta_0 + \theta_1 y_k + \theta_2 y_{k-1},$$

where $p_k(y_1, y_2, \dots, y_{k-1}, y_k)$ is the conditional probability of dropping out at time k , given the history of outcomes, $y_1, y_2, \dots, y_{k-1}$, the outcome subject to missingness, y_k , and the parameters θ_0 , θ_1 and θ_2 . For an MCAR process both θ_1 and θ_2 take the value zero, for an MAR process, θ_1 is zero and for an MNAR process all three of them are non-zeros. We simulated 500 data sets with SPlus (SPlus, 1999) and fit the mixed model, described earlier, assuming MCAR, MAR and MNAR processes using OSWALD (Smith et al., 1997). We note here that for the MNAR process OSWALD does not give standard errors of the parameter estimates. We estimated these standard errors from 200 bootstrap samples. As shown in Table 3, though the estimated parameters for the models and the corresponding standard errors were similar across the dropout process assumptions, the three selected performance criteria- (1) coverage (percent of time the true parameter is covered by the confidence interval), (2) percentage bias (relative magnitude of the raw bias to the standard errors), and (3) average width of the confidence intervals showed differences. For the group effect, the criterion of percentage bias favors the MNAR process. For the time effect two criteria, coverage and percentage bias, favor the MNAR process. For 100 simulated datasets, the ISNIs for the group effect were all above 1 and more than 80% of the ISNIs for the time effect were above 1. This suggests that, for those simulated datasets under an MNAR process, the analysis results are not really affected by assuming an MAR instead of an MNAR process, as reinforced by the results in Table 3.

9. Discussion

The increasing availability of software allows practitioners to analyze longitudinal psychiatric data with dropouts with more rigor than was previously feasible. Establishing treatment efficacy and durability is important in many psychiatric studies, and it relies on repeated assessment of study outcome variables. However, subjects may miss scheduled visits or drop

out of the study. The statistical literature is quite rich with methodological advances to account for dropout mechanisms in order to avoid bias in inference. However, there is a gap between the available theoretical developments and their practical implementation. In this paper, we have tried to bridge this gap. For simplicity, we have chosen to concentrate on continuous outcome measures and mixed effects models. The chief difficulties with discrete data concern the modeling and the computation of likelihood function. But they can be addressed similarly to the approach taken in the shared parameter model (Ten Have et al., 1998) or GEE (Robins et al., 1995). We have also reviewed methods that can be implemented using available software with minimal burden to the practitioner.

We emphasized the importance of collecting information regarding the reasons subjects drop out of a study or miss visits. Including this information as an integral part of study procedure can greatly facilitate the assessment of the nonignorability of missing data. We have shown how a sensitivity analysis assuming several dropout mechanisms, together with a critical look at the reasons for dropout, can help in judging whether it is necessary to resort to modeling specific nonignorable dropout processes (requiring computationally burdensome analyses). Our simulation study showed the potential impact of assumptions about the nature of the dropout processes. We showed that estimation of the uncertainty of the parameter estimates is likely to be affected if the assumed dropout process is in error. Pattern-mixture models are useful when patients with different missing-data patterns have quite different characteristics. We have provided an illustration for the applicability of the ISNI analysis. We agree with the developers of ISNI that it should be used as a practical tool to screen data sets for sensitivity to nonignorability.

We conclude by recommending first that recording the causes of dropouts should be protocolized in clinical trials. If an MCAR mechanism is suspected, then SAS Proc Mixed suffices for the analysis as if there were no missing values. If the mechanism is likely MAR, then OSWALD or an MI procedure that is consistent with the model for the repeated measure process produces valid inference. When potentially good predictors for the dropout mechanism are observed, analyzing the data as if it were MAR often works well though slight bias may be present theoretically. Otherwise efforts are needed to consider potential and most likely models for the dropout mechanism and a sensitivity analysis should be conducted afterwards. In the event of inconclusive results from a sensitivity analysis, further investigation of the dropout mechanism is warranted.

Acknowledgements

This research is supported in part by PHS grants P30 MH52247, P30 MH071944, P30 MH30915, K24 MH069340, R25MH60473, MH 073451, AG020677 and GlaxoSmithKline (who provided the pharmaceutical supplies for the conduct of the study reported).

Reference List

- Allison PD. Multiple imputation of missing data: A cautionary tale. *Sociological Methods and Research* 2000;28:301–309.
- Begley, AE.; Tang, G.; Mazumdar, S.; Houck, PR.; Scott, J.; Mulsant, BH.; Reynolds, CF. Computer Methods and Programs in Biomedicine. Department of Biostatistics, University of Pittsburgh; 2005. Use of OSWALD for analyzing longitudinal data with informative dropout. Technical Report #42
- Chatterjee, S.; Hadi, AS. Sensitivity Analysis in Linear Regression. New York: Wiley; 1988.
- Collins LM, Schafer JL, Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods* 2001;6:330–351. [PubMed: 11778676]
- Cook RD. Assessment of local influence. *Journal of the Royal Statistical Society, Series B* 1986;48:133–169.

- Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis. *Applied Statistics* 1994;43:49–93.
- Diggle, PJ.; Heagerty, P.; Liang, KY.; Zeger, SL. *Analysis of Longitudinal Data*. Oxford University Press; 2002.
- Fairclough, DL. *Design and Analysis of Quality of Life Studies in Clinical Trials*. Boca Raton, FL: CRC Press; 2002.
- First, MB.; Gibbon, M.; Spitzer, RL.; Williams, JBW. *Structured Clinical Interview for DSM-IV Axis I Disorders SCID I: Clinician version, administration booklet*. Washington, D.C.: American Psychiatric Press; 1997.
- Folstein MF, Folstein SW, McHugh PR. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 1975;12:189–198. [PubMed: 1202204]
- Gibbons RD, Hedeker D, Elkin I, Waternaux C, Kraemer HC, Greenhouse JB, Shea MT, Imber SD, Sotsky SM, Watkins JT. Some conceptual and statistical issues in analysis of longitudinal psychiatric data. Application to the NIMH treatment of Depression Collaborative Research Program dataset. *Archives of General Psychiatry* 1993;50:739–750. [PubMed: 8357299]
- Gueorguieva R, Krystal JH. Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the *Archives of General Psychiatry*. *Archives of General Psychiatry* 2004;61:310–317. [PubMed: 14993119]
- Hamilton M. A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry* 1960;23:56–62.
- Hedeker D. A mixed-effects multinomial logistic regression model. *Statistics in Medicine* 2003;22:1433–1446. [PubMed: 12704607]
- Hedeker D, Gibbons RD. Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods* 1997;2:64–78.
- Hogan JW, Roy J, Korkontzelou C. Tutorial in Biostatistics: Handling drop-out in longitudinal studies. *Statistics in Medicine* 2004;23:1455–1497. [PubMed: 15116353]
- Horton NJ, Lipsitz SR. Multiple imputation in practice: Comparison of software packages for regression models with Missing Variables. *The American Statistician* 2001;55:244–254.
- Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *Journal of American Statistical Association* 1952;47:663–685.
- Houck PR, Mazumdar S, Koru-Sengul T, Tang G, Mulsant BH, Pollock BG, Reynolds CF. Estimating treatment effects from longitudinal clinical trial data with missing values: Comparative analyses using different methods. *Psychiatry Research* 2004;129:209–215. [PubMed: 15590048]
- Kenward MG, Molenberghs G. Likelihood based frequentist inference when data are missing at random. *Statistical Science* 1998;13(No 3):236–247.
- Lavori PW, Dawson R, Shera D. A multiple imputation strategy for clinical trials with truncation of patient data. *Statistics in Medicine* 1995;14:1913–1925. [PubMed: 8532984]
- Little RJA. A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association* 1988;83:1198–1202.
- Little RJA. Modeling the drop-out mechanism in repeated measures studies. *Journal of American Statistical Association* 1995;90:1112–1121.
- Little, RJA.; Rubin, DB. *Statistical Analysis With Missing Data*. New York: Wiley; 1987.
- Ma G, Troxel AB, Heitjan DF. An index of local sensitivity to nonignorable drop-out in longitudinal modeling. *Statistics in Medicine* 2005;24:2129–2150. [PubMed: 15909292]
- Mallinckrodt CH, Sanger TM, Dube S, DeBrota DJ, Molenberghs G, Carroll RJ, Potter WZ, Tollefson GD. Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. *Biological Psychiatry* 2003;53:754–760. [PubMed: 12706959]
- Mazumdar S, Houck PR, Liu KS, Mulsant BH, Pollock BG, Dew MA, Reynolds CF. Intent-to-treat analysis for clinical trials: use of data collected after termination of treatment protocol. *Journal of Psychiatric Research* 2002;36:153–164. [PubMed: 11886693]
- Molenberghs G, Thijs H, Jansen I, Beunckens C, Kenward MG, Mallinckrodt C, Carroll RJ. Analyzing incomplete longitudinal clinical trial data. *Biostatistics* 2004;5:445–464. [PubMed: 15208205]

- Mulsant BH, Pollock BG, Nebes R, Miller MD, Sweet RA, Stack J, Houck PR, Bensasi S, Mazumdar S, Reynolds CF. A twelve-week double-blind randomized comparison of nortriptyline and paroxetine in older depressed inpatients and outpatients. *American Journal of Geriatric Psychiatry* 2001;9:406–414. [PubMed: 11739067]
- Robins JM, Rotnitzky A, Zhao LP. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of American Statistical Association* 1995;90:106–121.
- Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *Journal of American Statistical Association* 1984;79:516–524.
- Rubin DB. Inference and missing data. *Biometrika* 1976;63(No 3):581–592.
- Saltelli, A.; Chan, K.; Scott, EM. *Sensitivity Analysis*. John Wiley & Sons, Inc.; 2002.
- SAS Institute I. *SAS/STAT Software Version 9*. Cary, NC: SAS Institute, Inc; 2004.
- Schafer, JL. *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall; 1997.
- Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychological Methods* 2002;7:147–177. [PubMed: 12090408]
- Shen C, Weissfeld L. A copula model for repeated measurements with non-ignorable non-monotone missing outcome. *Statistics in Medicine*. 2005a
- Shen C, Weissfeld L. Application of pattern-mixture models to outcomes that are potentially missing not at random using pseudo maximum likelihood estimation. *Biostatistics* 2005b;6:333–347. [PubMed: 15772110]
- Smith, DM.; Robertson, WH.; Diggle, PJ. *Oswald: Object-oriented software for the analysis of longitudinal data in S*. Technical Report MA96/192. Lancaster University, Department of Mathematics and Statistics; 1997. <http://www.maths.lancs.ac.uk/Software/Oswald/> Lancaster University, Department of Mathematics and Statistics
- SPlus. *SPlus 2000 User's Guide*. Seattle: Mathsoft; 1999.
- Tang G, Little RJA, Raghunathan TE. Analysis of multivariate missing data with nonignorable nonresponse. *Biometrika* 2003;90:447–464.
- Tang, G.; Little, RJA.; Raghunathan, TE. Analysis of multivariate monotone missing data by a pseudolikelihood method. In: Lin, DY.; Heagerty, PJ., editors. *Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlative Data*. New York: Springer-Verlag; 2004.
- Ten Have TR, Kunselman AR, Pulkstenis EP, Landis JR. Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics* 1998;54:367–383. [PubMed: 9544529]
- Troxel AB, Ma G, Heitjan DF. An index of local sensitivity to nonignorability. *Statistical Sinica* 2004;14:1221–1237.
- Verbeke, G.; Molenbergh, G. *Linear Mixed Models for Longitudinal Data*. New York: Springer; 2000.
- Wu MC, Bailey KR. Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics* 1989;45:939–955. [PubMed: 2486189]
- Wu MC, Carroll RJ. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* 1988;44:175–188.
- Yuan, YC. Imputation for missing data: Concepts and new development; *Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference*; Cary: SAS Institute, Inc.; 2000. p. 1410-1419.
- Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986;42:121–130. [PubMed: 3719049]

Figure 1.

Flowchart for a general framework for the analysis of longitudinal data with dropouts

Nortriptyline N's = 54 52 49 44 42 41 37 35 31 30 27 24 24
 Paroxetine N's = 62 59 55 49 48 43 44 42 39 35 35 35 32

Figure 2.
 Mean Weekly 17-item Hamilton Rating Scale for Depression (HRSD) Profiles for Patients
 Randomized to Either Nortriptyline or Paroxetine.

Table 1
Results from the sensitivity analysis of the clinical trial data: parameter estimates and standard errors in parenthesis^{a,b,c}

Parameters	Selection Model MAR and ML Method	Selection Model MAR and ML Method	Selection Model MNAR and Logistic Dropout Model	Pattern- Mixture MAR and ML Method	Semi- Parametric Regression Model
Intercept	19.224 (0.674)	18.671 (0.691)	20.355 (0.604)	19.279 (0.656)	16.892 (0.836)
Group	0.195 (0.922)	0.213 (0.922)	0.119 (0.973)	0.284 (0.920)	1.680 (1.255)
Time	-1.308 (0.099)	-1.098 (0.086)	-1.145 (0.106)	-1.410 (0.115)	-0.927 (0.095)
Time × Group	0.089 (0.132)	0.085 (0.118)	0.099 (0.120)	0.190 (0.156)	-0.212 (0.141)

^a Mixed model for the outcome process

^b Logistic dropout model for the MNAR process (see text).

^c For definitions of MAR, MNAR, ML and MI we refer to the text.

*: P<0.05.

Table 2

Dropout reasons

	Nortriptyline N=54	Paroxetine N=62
Side Effects	18 (33%)	10 (16%)
Withdrew consent	4 (7%)	5 (8%)
Noncompliant with study procedures	1 (2%)	4 (6%)
Psychiatric reasons (onset of psychosis and delirium)	1 (2%)	3 (5%)
Medical reasons	1 (2%)	1 (2%)
Administrative (include study design)	2 (4%)	2 (3%)

Table 3

Results from the sensitivity analysis of the simulation study: parameter estimates, standard errors and performance criteria^a

(True parameters: Intercept= 20, Group = 1, Time = - .5)

	MCAR	MAR	MNAR					
Intercept								
Estimate	19.88	19.88	20.05					
Standard error	0.24	0.26	0.28					
Coverage (%)	93.8	93.0	92.2					
Bias (%)	0.60	0.60	0.25					
	0.94	1.02	1.10					
Width of Coverage								
Group								
Estimate	1.09	1.09	0.95					
Standard error	0.31	0.34	0.34					
Coverage (%)	94.0	96.2	93.0					
Bias (%)	9.00	9.00	5.00					
	1.22	1.33	1.33					
Width of Coverage								
Time								
Estimate	-0.53	-0.53	-0.51					
Standard error	0.02	0.02	0.03					
Coverage (%)	80.2	71.4	96.6					
Bias (%)	6.00	6.00	2.0					
	0.08	0.08	0.12					
Width of Coverage								
%Missing	Grp	Week0	Week2	Week4	Week6	Week8	Week10	Week12
	1 :	0	44.12	53.29	57.31	59.45	60.49	61.01
	Grp	0	20.13	29.72	34.53	37.91	39.84	41.20
	2 :							

^aFor definitions of MAR, MNAR, ML and MI we refer to the text.