BIOMETRICS 52, 1324–1333
December 1996

Intent-to-Treat Analysis for Longitudinal Studies with Drop-Outs

Roderick Little and Linda Yau

Department of Biostatistics, University of Michigan,
Ann Arbor, Michigan 48109, U.S.A.

SUMMARY

We consider intent-to-treat (IT) analysis of clinical trials involving longitudinal data subject to drop-out. Common methods, such as Last Observation Carried Forward imputation or incomplete-data methods based on models that assume random dropout, have serious drawbacks in the IT setting. We propose a method that involves multiple imputation of the missing values following drop-out based on an "as treated" model, using actual dose after drop-out if this is known, or imputed doses that incorporate a variety of plausible alternative assumptions if unknown. The multiply-imputed data sets are then analyzed using IT methods, where subjects are classified by randomization group rather than by the dose actually received. Results from the multiply-imputed data sets are combined using the methods of Rubin (1987, *Multiple Imputation for Nonresponse in Surveys*). A novel feature of the proposed method is that the models for imputation differ from the model used for the analysis of the filled-in data. The method is applied to data on a clinical trial for Tacrine in the treatment of Alzheimer's disease.

1. Introduction

Intent-to-treat (IT) analysis in clinical trials embodies two principles: all individuals randomized to a treatment arm should be included in the analysis, including those who drop out prematurely; and treatment effects should be measured with subjects assigned to the treatment to which an individual is randomized ("as randomized"), rather than to the treatment actually received ("as treated"). Thus, subjects randomized to a control treatment who obtain and receive the new treatment are regarded as controls rather than treated cases, and subjects who fail to comply with a new treatment are analyzed as if the treatment was received.

IT analysis is controversial. Analyses based on actual treatment received are arguably more relevant to an "explanatory" analysis (Schwartz and Lellouch, 1967) that concerns the biological effects of the treatment, but they are prone to selection bias when subjects take a different treatment from that assigned to them. The IT, or pragmatic analysis in the terminology of Schwartz and Lellouch (1967), is less prone to bias in estimating its objective, namely the effect of treatment assignment, particularly if treatments are randomly assigned. Also, the effect of treatment assignment provides information about the practical utility of a treatment. If, for example, a potent and effective drug is avoided by the majority of subjects because of unpleasant side effects, then this negative feature is taken into account in the IT analysis, but it is ignored in an "as treated" analysis focused on effects for those who took the treatment. On a practical level, IT analysis is often one of the set of analyses required by agencies that assess and approve new medications. Our objective is to discuss analyses of the IT type for longitudinal studies with drop-outs, and propose a new approach based on multiple imputation which we believe meets the objectives of IT analysis in a more satisfactory way than methods in current use.

We consider the analysis of randomized clinical trials, where (possibly after a run-in period) subjects are randomized to one of K treatments, and then data are collected at T time points up until the conclusion of the study. Summary measures of outcome are then computed and treatments

Key words: Clinical trials; Imputation; Incomplete data; Missing values; Multiple imputation; Repeated measures.

are compared. The following example of an important clinical trial is presented to help focus and illustrate issues.

Example 1. Drop-outs in a dose-titration study of Tacrine for Alzheimer's disease. A double-blind, placebo controlled, parallel group study was conducted for the treatment of Alzheimer's disease using the drug Tacrine, also known as Cognex (Knapp et al., 1994). After establishing eligibility and completing baseline assessments, patients were randomized to control and treatment groups, and a variety of cognitive measures were administered periodically up to 18 weeks from baseline (a second set of outcomes measured through 30 weeks after baseline is not considered here). For the 18-week study, subjects were randomized to three treatment groups: a placebo group; a low-dose group where patients were given a 40 mg/day dose of Tacrine for 6 weeks and an 80 mg/day dose for weeks 7–18; and a high-dose group where patients were given a 40 mg/day dose for 6 weeks, an 80 mg/day dose for weeks 7–12, and a 120 mg/day dose for weeks 13–18. The titration to higher doses of Tacrine was used because potentially serious side effects on liver function dictated against initial administration of a high dose of the drug. Measures were taken at baseline and after weeks 6, 12, 16, and 18. We will focus here on data for a primary outcome measure called the Alzheimers' Disease Assessment Scale, Cognitive component (ADAS-Cog) (see Rosen, Mohs, and Davis (1984) for more information on this scale).

For the purposes of this article, we define drop-outs to be subjects who were no longer part of the double-blind study. The study analysis was complicated by a substantial number of dose-related drop-outs primarily due either to protocol rules, specifically an elevated liver function test, or to other reasons, such as lack of compliance or various dose-related adverse events (Watkins et al., 1994). In particular, at week 18, 44 of the 184 subjects (24%) in the placebo group, 31 of the 61 subjects (51%) in the low-dose group, and 244 of the 418 subjects (57%) in the high-dose group were no longer in the double-blind phase of the study. For simplicity, we confine attention here to the complete cases and drop-outs who form a monotone missing-data pattern, omitting a small number of subjects who missed a visit, but reentered the study double-blind after missing a visit. The method we propose can be extended to handle nonmonotone patterns. An analysis of the cases remaining in the study at week 18 suggested a significant dose-related effect for Tacrine, but the high and differential drop-out rates were a cause for concern. The study protocol included as one of the required analyses an intent-to-treat analysis of the data at week 18, using analysis of covariance adjusted for the baseline ADAS-Cog values.

IT principles dictate that all cases, complete or incomplete, should be included in the analysis. More specifically, we regard the target estimand of IT analysis as the estimand of the "as randomized" analysis of the hypothetical complete data, if outcomes were available for the subjects that dropped out. Since the nature of those outcomes depend on information about behavior and dose after drop-out that is often not available, assumptions need to be made. Thus, for longitudinal data with drop-outs, IT analysis is often not uniquely defined, and the correct definition of the analysis may be debatable. Given lack of consensus on this point, we favor a sensitivity analysis that displays the effects of alternative assumptions about dosage behavior after drop-out. We describe such an analysis for the Tacrine data set.

Given that our objective is to mimic a complete-data analysis, IT analysis requires a method of dealing with the missing data. In the Tacrine trial, a primary outcome measure is the change in ADAS-Cog score between baseline and week 18, but the week 18 score is not available for all cases. Attempts were made to follow drop-outs, and week 18 ADAS-Cog scores were measured for a subset, called the "retrieved drop-outs." While all our IT analyses include these subjects, it is not clear whether information retrieved after drop-out should be included in the analysis since (a) subjects were no longer blinded to treatment and (b) dosages of Tacrine administered after drop-out did not follow the treatment protocol. For example, some control subjects received Tacrine after dropout. Including the observed outcomes for such cases clearly distorts the treatment–control comparison.

A referee suggests that data on retrieved drop-outs may be used as a basis for imputing values for nonretrieved drop-outs. If in a particular application (a) and (b) are not considered important problems, and hence information about the retrieved drop-outs is considered reliable, then this would indeed be a sensible approach. However, it is not an option that is pursued in the Tacrine application.

In any event, a value for week 18 needs to be imputed for the unretrieved (and arguably also for the retrieved) drop-outs or the incomplete cases need to be included in some other fashion. The treatment of the incomplete data for the IT analysis is the main focus of this article.

2. Current Approaches

2.1 *Last Observation Carried Forward Imputation*

Heyting, Tolboom, and Essers (1992) provide a useful review of current methods for handling drop-outs in longitudinal clinical trials. The common practice of confining analysis to the complete cases is not satisfactory for IT analysis, as noted above. A popular imputation procedure is Last Observation Carried Forward (LOCF), which simply fills in missing values for an incomplete case by the last observed value for that case (see, for example, Pocock (1983, Section 12.3)). The LOCF method assumes, in effect, that outcome remains constant at the last observed value after drop-out, which seems unlikely in many applications. If the treatment induces a change in the outcome and otherwise the outcome is stable, then LOCF may be somewhat plausible for drop-outs deprived the effect of the treatment. To the extent that the prior treatments have carry-over effects after drop-out, the LOCF assumption might be regarded as conservative for estimating the size of treatment effects. However, LOCF is likely to be anti-conservative in the Tacrine application. With no treatment, Alzheimer's disease patients generally display a steady downward course for cognitive measures such as ADAS-Cog. If LOCF is applied and drop-outs concentrate in the treatment arms, an ineffective treatment may be incorrectly found to be effective because LOCF interrupts a deteriorating course after drop-out. Any imputation procedure incorporates an assumption about the predictive distribution of the missing values given the observed data, and should not be adopted unless this underlying assumption is reasonable.

2.1 *IT Repeated-Measures Models*

An alternative approach to imputation is to apply a method that analyzes the observed data as incomplete, using methods such as maximum likelihood (ML) or generalized estimating equations (GEE). In particular, ML for normal outcomes can be based on the model discussed in Laird and Ware (1982) and Jennrich and Schluchter (1986), which allows a flexible choice of mean and covariance structure. Versions of this model are implemented in a number of statistical software packages, including BMDP (BMDP5V; see Dixon, 1988) and SAS (Proc Mixed; see SAS, 1992). For discussions of the ML approach in clinical trial settings, see Murray and Findlay (1988) and associated correspondence (Harris, 1993) or Heyting et al. (1992). See Smith (1996) for an "as treated" ML analysis of the Tacrine data.

Example 2. Example 1 continued. The first two rows of Table 1 show results of two analyses based on this approach. Let $y_i = (y_{i1}, y_{i2}, \ldots, y_{iT})$ denote the complete vector of repeated measures for subject i. In our application, $T = 4$ and $y_{it} = x_{it} - x_{i0}$, where x_{it} is the ADAS-Cog score for subject i at time t and x_{i0} is the ADAS-Cog score at baseline time 0. Thus the outcomes are changes in ADAS-Cog score from baseline at weeks 6, 12, 16, and 18. The distribution of y_i, given treatment randomization group indicators r_i and other covariates z_i, is modeled as multivariate normal with mean ν_i and covariance matrix Σ, where

$$\nu_{it} = \mathrm{E}(y_{it}|r_i, z_i) = \mu_{it}^{(0)} + \delta_t^T r_i. \tag{2.1}$$

Here $\mu_{it}^{(0)}$ represents the mean of y_{it} for subjects in the reference (control) group, and r_i represents a $(K-1) \times 1$ vector of randomization group indicators for subject i. In the Tacrine application $K = 3$, r_{i1} takes the value 1 for an individual randomized to the low-dose group and 0 otherwise, r_{i2} takes the value 1 for an individual randomized to the high-dose group and 0 otherwise, and $\delta_t^T = (\delta_{t1}, \delta_{t2})$ represent IT treatment effects. The means $\mu_{it}^{(0)}$ are modeled as linear functions of other covariates z_i. Specifically,

$$\mu_{it}^{(0)} = \beta_0 + \tau_t + \lambda_{s(i),t} + \alpha_t x_{i0}, \tag{2.2}$$

where τ_t represents a time effect, $s(i)$ is the treatment site for subject i so that $\{\lambda_{s(i),t}\}$ represent site effects, and α_t is a linear regression coefficient on the baseline ADAS-Cog value.

The row labeled (1A) in Table 1 shows estimates of $(\delta_{41}, \delta_{42})$, the low- and high-dose treatment effects at week 18, and associated asymptotic standard errors when this model is fitted to the complete cases, discarding drop-outs. The model was fitted by restricted maximum likelihood using SAS Proc Mixed, and assumes normal outcomes and an unstructured covariance matrix. The results suggest significant effects of Tacrine for both dosage levels, but with smaller effects for the higher dosage group. The row labeled (1B) shows the results from fitting the same model to all cases lying within the monotone data pattern, that is, the complete cases and the drop-outs. The estimated treatment effect for the high-dose group is increased when the incomplete cases are in-

Table 1

Summary of analyses of week 18 Tacrine data

Analysis	Multiple imputation	80 mg/day			120 mg/day		
		Treat. diff. (s.e.)	p-value	C.I.	Treat. diff. (s.e.)	p-value	C.I.
(I) Intent-to-treat restricted ML estimates based on observed data, ignorable drop-out, unstructured covariance matrix							
(1A) Complete cases		−3.584 (1.287)	0.0054	(−6.11, −1.06)	−2.236 (0.705)	0.0016	(−3.62, −0.85)
(1B) Monotone pattern		−3.673 (1.208)	0.0024	(−6.04, −1.30)	−3.436 (0.667)	0.0000	(−4.74, −2.13)
(II) Intent-to-treat analyses of data multiply-imputed under alternative imputation models							
(2A) Continuing-dose	1	−3.486 (0.951)	0.0003	(−5.35, −1.62)	−4.146 (0.563)	0.0000	(−5.25, −3.04)
	2	−3.682 (0.876)	0.0000	(−5.40, −1.97)	−3.101 (0.518)	0.0000	(−4.12, −2.08)
	3	−3.142 (0.944)	0.0009	(−4.99, −1.29)	−3.567 (0.559)	0.0000	(−4.66, −2.47)
	4	−4.889 (0.908)	0.0000	(−6.67, −3.11)	−3.011 (0.538)	0.0000	(−4.07, −1.96)
	5	−4.633 (0.910)	0.0000	(−6.42, −2.85)	−3.395 (0.538)	0.0000	(−4.45, −2.34)
	6	−4.146 (0.920)	0.0000	(−5.95, −2.34)	−3.642 (0.545)	0.0000	(−4.71, −2.57)
	7	−5.239 (0.925)	0.0000	(−7.05, −3.43)	−3.255 (0.547)	0.0000	(−4.33, −2.18)
	8	−4.463 (0.933)	0.0000	(−6.29, −2.63)	−3.695 (0.552)	0.0000	(−4.78, −2.61)
	9	−4.511 (0.953)	0.0000	(−6.38, −2.64)	−3.762 (0.564)	0.0000	(−4.87, −2.66)
	10	−3.497 (0.899)	0.0001	(−5.26, −1.73)	−3.104 (0.532)	0.0000	(−4.15, −2.06)
	MI inference	−4.169 (1.173)	0.0039	(−6.72, −1.62)	−3.468 (0.663)	0.0002	(−4.90, −2.04)
(3A) Zero-dose	MI inference	−2.097 (1.087)	0.0734	(−4.42, 0.23)	−1.829 (0.688)	0.0203	(−3.32,−0.34)
(4A) Nearest-dose	MI inference	−2.787 (1.097)	0.0230	(−5.13, −0.44)	−2.727 (0.656)	0.0010	(−4.13, −1.32)
(2B) Continuing-dose incl. follow-ups	MI inference	−3.349 (1.020)	0.0032	(−5.46, −1.24)	−2.044 (0.597)	0.0020	(−3.27, −0.82)
(3B) Zero-dose incl. follow-ups	MI inference	−2.466 (1.024)	0.0248	(−4.59, −0.34)	−1.489 (0.598)	0.0197	(−2.72, −0.26)
(4B) Nearest-dose incl. follow-ups	MI inference	−2.658 (1.020)	0.0156	(−4.76, −0.55)	−1.649 (0.594)	0.0094	(−2.86, −0.44)
(3C) Zero-dose without adjustment	MI inference	−1.906 (1.029)	0.0828	(−4.09, 0.28)	−0.941 (0.609)	0.1422	(−2.23, 0.35)

cluded, and is now similar to the estimate for the low-dose group. Both treatment effects are highly statistically significant.

The estimates in the second row make the key assumption that the drop-out mechanism is missing at random (Rubin, 1976; Little and Rubin, 1987), which means that drop-out at time t does not depend on unobserved outcomes at times $t' \geq t$, after conditioning on the observed data up to time t included in the model. Diggle and Kenward (1994) use the term random dropout (RD) for this assumption. The propensity weighting approach discussed in Heyting et al. (1992) also makes this assumption.

The RD assumption is often sensible, but it is highly questionable in our IT analysis setting. The IT analysis conditions on the treatment group to which the subject was randomized, specifically dummies r_i for control, low- and high-dose groups in the Tacrine example, but the analysis does not condition on the actual treatments received, specifically the dosages of Tacrine administered. Since the original randomization group is unchanged by drop-out, the IT analysis of all the observed data based on the RD assumption does not model any changes in dose after drop-out, even when we know that they occurred. In the Tacrine trial, subjects in the high-dose group who dropped out of the study because of liver enzyme elevations received reduced or zero doses after drop-out, but the RD analysis treats them as if they still received the high-dose treatment, thus potentially overstating the intent-to-treat effect associated with the drug. Thus, a different analysis is needed.

Models for nonrandom dropout have been considered (see, for example, Wu and Carroll (1988); Wu and Bailey (1989); Diggle and Kenward (1994); Little and Wang (1996); Little (1995)). However, estimation of these models is highly sensitive to model misspecification (Little, 1995). A better approach, in our view, is to impute values of the outcome after drop-out using models for the missing data that condition on all relevant observed data, and in particular on information about treatments actually administered (as opposed to randomized). Imputation uncertainty is handled by multiple imputation (Rubin, 1987; Rubin and Schenker, 1991), where $M > 1$ sets of imputes are created for the missing values in the data set, as draws from the predictive distribution of the missing values under an assumed model. An IT model such as that in equations (2.1) and (2.2) is then fitted to each of the M filled-in data sets. The results are combined using the multiple imputation technology discussed in Rubin (1987, Chapter 3).

Specifically, given large samples, inference for a scalar parameter θ with complete data is typically based on a point estimate $\hat{\theta}$, a variance U, and a normal reference distribution; for example, the typical interval estimate has the form $\hat{\theta} \pm z_{(1-\alpha/2)} U^{1/2}$, where $z_{(1-\alpha/2)}$ is the $1-\alpha/2$ quantile of the standard normal distribution. In the presence of nonresponse, M imputations of the missing values are created under the posited model for the missing data, yielding M completed data sets and hence M sets of completed-data statistics, say $\hat{\theta}_k$ and U_k, $k = 1, \ldots, M$.

The M sets of completed-data statistics are combined to create one multiple-imputation inference as follows. The estimate of θ is $\bar{\theta} = M^{-1} \Sigma_{k=1}^{M} \hat{\theta}_k$,the average of the M completed-data estimates of θ. Also, let $\bar{U} = M^{-1} \Sigma_{k=1}^{M} U_k$ be the average of the M completed-data variances, and $B = (M-1)^{-1} \sum_{k=1}^{M} (\hat{\theta}_k - \bar{\theta})^2$ be the between-imputation variance of the completed-data estimates of θ. Then the total variance of $(\theta - \bar{\theta})$ is given by the sum of the within-imputation component $(\bar{U})$ and the between-imputation component (B) multiplied by a finite M correction $(1 + M^{-1})$, that is $V = \bar{U} + (1 + M^{-1})B$. Interval estimates and significance levels are obtained using a t distribution with center $\bar{\theta}$, scale $V^{1/2}$, and degrees of freedom $\nu = (M-1)(1+r^{-1})^2$, where $r = (1+M^{-1})B/\bar{U}$ is the ratio of the between-imputation component of variance to the within-imputation component. Thus, for example, a $100(1-\alpha)\%$ interval estimate for θ is $\bar{\theta} \pm t_{\nu,(1-\alpha/2)} V^{1/2}$, where $t_{\nu,(1-\alpha/2)}$ is the $1-\alpha/2$ quantile of the t-distribution with ν degrees of freedom. In general, r estimates the quantity $\gamma/(1-\gamma)$, where γ is the fraction of information about θ that is missing due to nonresponse. The two-sided p-value for the null hypothesis H_0: $\theta = 0$ is computed by comparing $\bar{\theta}/V^{1/2}$ with a t-distribution with ν degrees of freedom.

3. The Imputation Model

Our imputation model conditions on the treatments actually administered, and thus does require knowledge of these treatments after drop-out. The question of how to proceed when this information is not available is deferred until the following section. The following as-treated imputation model was applied to the Tacrine data set. As before, $y_i = (y_{i1}, y_{i2}, \ldots, y_{iT})$ denotes the vector of repeated measures for subject i; in our application $T = 4$ and y_{it} represents the change in ADAS-Cog score between time t and baseline. The distribution of y_i given r_i, z_i, and actual dosage indicators d_i is

modeled as the product of the following conditional normal regression models:

$$(y_{it} \mid y_{i0}, y_{i1}, \ldots, y_{i,t-1}, r_i, d_i, z_i) \overset{\sim}{} {}_{ind}\mathrm{N}(\mu_{it}, \sigma_{tt\cdot 12\cdots t-1}), \qquad t = 1, \ldots, T, \tag{3.1}$$

where in the Tacrine application

$$\mu_{i1} = \beta_{01} + \alpha_1 x_{i0} + \lambda_{s(i),1} + \gamma_1^T d_{i1}, \tag{3.2}$$

$$\mu_{it} = \beta_{0t} + \alpha_t x_{i0} + \gamma_t^T d_{it} + \sum_{j=1}^{t-1} \beta_{tj} y_{ij}, \qquad t = 2, \ldots, 4. \tag{3.3}$$

Here d_{it} represents a vector of indicators for the actual treatment for subject i at time t. That is, d_{it1} takes the value 1 for an actual 40 mg dose at time t and 0 otherwise, d_{it2} takes the value 1 for an actual 80 mg dose at time t and 0 otherwise, d_{it3} takes the value 1 for an actual 120 mg dose at time t and 0 otherwise, and $\gamma_t = (\gamma_{1t}, \gamma_{2t}, \gamma_{3t})$ represents dose-specific treatment effects at time t. The coefficients $\{\beta_{tj}\}$ model the association of the outcome at time t with the values of outcomes at previous time points. The means are modeled as linear functions of the baseline score x_{i0} and site-specific effects, although the latter are included only for $t = 1$ because they added little predictive power to the regressions for later time points. The conditioning on r_i in equation (3.1) is included to make explicit an assumption that the distribution of y_i given r_i, d_i, and z_i does not depend on r_i, that is, outcomes depend on the treatments actually received rather than on the treatments randomized. This assumption is related to the "weak exclusion restriction of treatment assignment given treatment received" in Imbens and Rubin (1996).

This model for the set of conditional distributions implies that the joint distribution of y_i is multivariate normal with marginal mean ν_i and unstructured covariance matrix Σ, where ν_i has tth component

$$\nu_{it} = \mathrm{E}(y_{it} | r_i, d_i, z_i) = \mu_{it}^{(0)} + \sum_{j=1}^{t} \gamma_{tj}^{*T} d_{ij}. \tag{3.4}$$

The regression coefficients for doses in the marginal model are related to the coefficients in the conditional models by the expressions

$$\begin{aligned}
\gamma_{tt}^* &= \gamma_t, \qquad t = 1, \ldots, 4 \\
\gamma_{t,t-1}^* &= \beta_{t,t-1}\gamma_{t-1}, \qquad t = 2, \ldots, 4 \\
\gamma_{t,t-2}^* &= (\beta_{t,t-2} + \beta_{t,t-1}\beta_{t-1,t-2})\gamma_{t-2}, \qquad t = 3, 4 \\
\gamma_{41}^* &= (\beta_{41} + \beta_{43}\beta_{31} + \beta_{42}\beta_{21} + \beta_{43}\beta_{32}\beta_{21})\gamma_1.
\end{aligned}$$

An important feature of the marginal model is that the marginal mean at time t depends not just on the treatment at time t, but on the entire set of treatments up to and including time t. The coefficient γ_t^* represents the incremental effect of treatment at time t that is lost when a subject drops out at time t.

Multiple imputations based on this model were created sequentially, first filling in missing values of y_{i2} as draws from the predictive distribution of y_{i2} given y_{i1}, then filling in missing values of y_{i3} as draws from the predictive distribution of y_{i3} given observed or imputed values of y_{i1}, y_{i2}, and so on. To account for uncertainty in parameter estimates, model parameters are first drawn from their posterior distribution and then missing values drawn from their predictive distribution conditional on the drawn parameters. Here, regression coefficients arising in these predictions were drawn from their posterior distributions based on the available data up to that time point, as is appropriate for the monotone missing-data pattern. Specifically, a residual variance was drawn as the residual sum of squares divided by a chi-squared deviate with residual degrees of freedom, and then regression coefficients were drawn as multivariate normal centered at their estimates, with covariance matrix given by the inverse of the design matrix multiplied by the drawn residual variance (Rubin, 1987).

4. Assumptions About Dose After Drop-Out

Since the imputation model conditions on the treatment information d_i, it requires assumptions about the nature of the treatment after drop-out. In some studies, information about dose after drop-out may be unknown or unreliable. In the Tacrine study, an attempt was made to follow cases after drop-out, and dosage information was recorded. Dosage information might be used for imputations, although it could be argued that this limits the generalizability of the study to cases

where similar dosage patterns are predicted in the real population; for example, it is questionable whether nonzero doses for Tacrine should be allowed when imputing for control cases, even if they actually occurred. Since there seems no obvious single answer to these questions, we carry out a sensitivity analysis of the Tacrine data set, where imputations are obtained for a range of alternative assumptions about dose after drop-out.

The first assumption is that the subject continues on the same treatment as that immediately prior to drop-out. We call this the "continuing-dose model." Cases in the high-dose groups that drop out prior to the protocol dose increase are imputed to receive that increase. While this model seems unrealistic in the Tacrine example when drop-out is caused by the side-effect, the analysis is interesting since it is the assumption implicit in an incomplete-data analysis based on an ignorable model for the drop-out process, such as would be obtained from programs such as BMDPAM or SAS Proc Mixed. It might be expected to yield optimistic estimates of IT treatment effects, if useful treatment effects exist and drop-outs in the treatment group in fact receive reduced doses after drop-out. Note that if treatment group "as randomized" is similar to treatment group "as treated," that is, there is good compliance with the randomized treatment, then imputation under the "continuing-dose" model should yield similar estimates to the corresponding IT analysis of the incomplete data assuming RD, that is, the analysis in row (1B) of Table 1.

Our second assumption is that the subject reverts to the control treatment after drop-out, that is, the placebo treatment in the Tacrine example. We call this the "zero-dose model." This analysis tends to minimize differences between treatment and control after drop-out, and hence might be expected to yield deflated estimates of IT treatment effects. It may be argued that the zero-dose model is too conservative here since there is evidence that the liver side effect of Tacrine is a transient effect, and many who dropped out from this side-effect in fact resumed Tacrine after a time. Analyses under the "continuing-dose" and "zero-dose" models plausibly bracket the actual effects of the treatments in this and other situations.

Our final assumption is the "nearest-dose" model, where (a) cases in the control group are assigned a zero dose after drop-out and (b) cases in the treatment groups are assigned a treatment-group dose (0, 40, 80, or 120 mg) that is closest to the actual recorded dose after drop-out, rounding up in the case of ties. The actual dose cannot be used here since the imputation model treats dosage as a categorical rather than a continuous variable. This model seems the most realistic for the Tacrine example, although it is "counter factual" in that some control cases took Tacrine after drop-out.

Example 3. Example 2 continued. Table 1 shows IT analyses of the data set imputed under alternative assumptions about dose after drop-out. Note that the IT analysis is the same in all these analyses; it is the multiple imputations that change.

Block (2A) shows the results of fitting the IT model of equations (2.1) and (2.2) to $M = 10$ data sets multiply-imputed under the "continuing-dose" model. The multiple-imputation inference, summarized in the "MI inference" row, shows estimates, standard errors, p-values, and confidence intervals computed using the methods summarized in Section 2.2. The results are quite similar to the results in row (1B) for the ignorable IT model fitted to the monotone incomplete data, illustrating that these two analyses make similar assumptions. Ten multiply-imputed data sets were also generated for all the other imputation models in Table 1 and analyzed in the same way. Results for each of the individual imputed data sets are omitted for these models to save space.

Block (3A) shows the results of fitting the IT model of equations (2.1) and (2.2) to data sets multiply-imputed under the "zero-dose" model. Note that the summary estimated treatment effects are half the size of the estimates under the "continuing-dose" model (-2.10 versus -4.17 for 80 mg and -1.83 versus -3.47 for 120 mg). The effect for the 120 mg dose remains significant at the 5% level ($p = .02$), but the effect for the 80 mg dose no longer achieves statistical significance at the 5% level ($p = .07$). These results illustrate the more conservative nature of the IT analysis based on the "zero-dose" model for this application.

Estimated treatment effects for the "nearest-dose" model are shown in block (4A), and are -2.79 for 80 mg ($p = .02$) and -2.73 for 120 mg ($p = .001$). These estimates, arguably the most realistic for the Tacrine data set, lie between those for the "zero-dose" and "continuing-dose" models, reflecting predicted effects of reduced but nonzero Tacrine doses after drop-out.

The results in blocks (2A) to (4A) do not make use of outcome data at week 18 for the retrieved drop-outs. The analyses in blocks (2B), (3B), and (4B) parallel those in blocks (2A), (3A) and (4A), but with outcome data from the retrieved drop-outs at week 18 included. Results are less sensitive to the choice of imputation model and standard errors are lower, reflecting the fact that

less data are multiply-imputed. Estimated treatment effects are a bit lower than for the models that do not include outcome data for retrieved drop-outs, although they are statistically significant.

An important feature of the proposed analysis is that effects of treatment prior to drop-out are allowed to extend to effects after drop-out via the estimates of the coefficients β_{tj} in model (3.3). An excessively conservative analysis imputes the drop-outs in the treatment group like controls, ignoring carryover effects of previous treatments. This approach can be illustrated by fitting our imputation model with the coefficients β_{tj} set to zero. Results are presented in block (3C) of Table 1, labeled "zero-dose without adjustment." Note that the treatment effects are smaller than for the "zero-dose model" in block (3A), indicating that there is some predicted treatment gain for individuals who drop out before the end of the study and then receive a zero treatment dose.

An important assumption of the proposed imputation model is that the conditional distribution of outcomes after drop-out, given outcomes prior to drop-out, dose, and other covariate information, is the same as the conditional distribution for cases continuing in the study. In the Tacrine study, there is essentially no information about the effects on outcome of switching to lower doses, so this assumption cannot be checked empirically. However, some such assumption is needed to include the incomplete cases in the analysis, and the assumption adopted seems plausible.

5. Discussion

We propose a new form of IT analysis for longitudinal data with drop-outs, where values of the outcome after drop-out are multiply-imputed using a model that conditions on actual or assumed treatments received, and then a classical IT analysis based on the treatment as randomized is applied to the imputed data. The analysis was applied to a real example, producing a variety of estimates of treatment effects, depending on assumptions about dose after drop-out and the treatment of outcome data collected for retrieved drop-outs. The models presented here were formulated for the specific application, but the ideas underlying the analysis clearly apply much more generally.

The analysis of the Tacrine data reflects our belief that there is often no single definitive IT analysis for the repeated-measures problem considered. Any analysis requires assumptions about the dose after drop-out, and the choice of assumptions is subject to debate. Ideally, one should make assumptions that reflect how the drug will be used in the real world after the study, but knowledge about this will usually be sketchy at the clinical trial phase. Given this reality, an analysis that assesses sensitivity of answers to alternative assumptions, as in Table 1, may be the best option. The lack of a definitive IT analysis has design implications for longitudinal clinical trials where drop-outs are likely. The Tacrine trial attempted to collect information about dosages and outcomes after drop-out. While this information is often problematic because of its observational nature and the absence of blinding, it seems preferable to have it available for one or more of the sensitivity analyses.

Our analyses are based on multiple imputation of the missing data. It would be possible in principle to derive an analysis that does not require imputation of the missing data, by expressing parameters of the IT model as functions of parameters of the imputation model and then deriving ML inference for those functions. However, multiple imputation provides a simple way of generating the IT inference without the need for special software and programming. Indeed, a general advantage of multiple imputation is that differences between the imputation and analysis model are readily accommodated. Here the crucial difference is that the imputation model conditions on actual treatment (observed or imputed) rather than the treatment randomized. More generally, observed variables that improve the imputations should be included in the imputation model, even if they are not included in the final IT model because they are not considered exogenous to the treatment.

Programming of our methods was carried out using SAS macros (SAS, 1989). An annotated listing of the SAS macro for the Tacrine example is available in Dr. Little's World Wide Web home page: http://www.sph.umich.edu/~rlittle, and can be modified for other applications.

Acknowledgements

This work was supported by NSF grant DMS-9408837 and by NIMH grant P30 MH38330-12. While taking full responsibility for errors, we thank Wayne Ewy, Stephen Gracon, Don Rubin, two referees, and an associate editor for valuable comments on an earlier draft, and Parke Davis Pharmaceuticals for permission to use the Tacrine data set to illustrate our methods.

Résumé

Nous considérons l'analyse en intention de traitement d'essais cliniques comportant des données longitudinales susceptibles d'être incomplètes du fait de sorties d'étude. Les méthodes habituelles, telles que la propongation de la dernière valeur observée ou les méthodes pour données incomplètes construites à partir de modèles qui supposent des sorties d'étude eléatoires, sont sujettes à caution dans le contexte de l'intention de traitement. La méthode que nous proposons est une imputation multiple des données manquantes après une sortie d'étude, dans le cadre d'un modèle qui tient compte du traitement réellement reçu (il s'agit de la dose réellement reçue après la sortie d'étude si cette dose est connue, ou des doses imputées—selon différents scénarios plausibles—dans le cas où cette dose est inconnue). Les données ainsi complétées sont ensuite analysées par une approche en intention de traitement où les sujets sont affectés à leur groupe de randomisation sans tenir compte de la dose qu'ils ont réellement reçue. Les résultats obtenus à partir de ces données complétées par imputations multiples sont combinés selon les méthodes de Rubin (1987). Une caractéristique nouvelle de la méthode proposée est que les modèles utilisés pour imputer les données manquantes sont distincts de ceux utilisés pour analyser les données ainsi complétées. Cette méthode est appliquée aux données d'un essai clinique de la Tacrine dans la maladie d'Alzheimer.

References

Diggle, P. and Kenward, M. G. (1994). Informative dropout in longitudinal data analysis (with Discussion). *Applied Statistics* **43,** 49–94.

Dixon, W. J. (1988). *BMDP Statistical Software.* Berkeley: University of California Press.

Harris, R. I. (1993). Statistical handling of drop-outs in longitudinal clinical trials (Letter: Comment). *Statistics in Medicine* **12**(23), 2248–2250.

Heyting, A., Tolboom, J. T. B. M., and Essers, J. G. A. (1992). Statistical handling of drop-outs in longitudinal clinical trials. *Statistics in Medicine* **11,** 2043–2063.

Imbens, G. W. and Rubin, D. B. (1996). Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics*, in press.

Jennrich, R. I. and Schluchter, M. D. (1986). Incomplete repeated-measures models with structured covariance matrices. *Biometrics* **42,** 805–820.

Knapp, M. J., Knopman, D. S., Solomon, P. R., Pendlebury, W. W., Davis, C. S., and Gracon, S. I. (1994). A 30-week randomized controlled trial of high-dose Tacrine in patients with Alzheimer's disease. *Journal of the American Medical Association* **271**(13), 985–991.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38,** 963–974.

Little, R. J. A. (1995). Modeling the drop-out mechanism in longitudinal studies. *Journal of the American Statistical Association* **90,** 1112–1121.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* New York: Wiley.

Little, R. J. A. and Wang, Y. (1996). Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics*, **52,** 98–111.

Murray, G. D. and Findlay, J. G. (1988). Correcting for the bias caused by drop-outs in hypertension trials. *Statistics in Medicine* **7,** 941–946.

Pocock, S. J. (1983). *Clinical Trials: A Practical Approach.* New York: Wiley.

Rosen, W. G., Mohs, R. C., and Davis, K. L. (1984). A new rating scale for Alzheimer's disease. *American Journal of Psychiatry* **141,** 1356–1364.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63,** 581–592.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.

Rubin, D. B. and Schenker, N. (1991). Multiple imputation in health-care databases: An overview and some applications. *Statistics in Medicine* **10,** 585–598.

SAS. (1989). *SAS Guide to Macro Processing,* Version 6, 2nd edition. Cary, North Carolina: SAS Institute, Inc.

SAS. (1992). The mixed procedure. In *SAS/STAT Software: Changes and Enhancements,* Release 6.07, Technical Report P-229, Chapter 16. Cary, North Carolina: SAS Institute, Inc.

Schwartz, D. and Lellouch, J. (1967). Explanatory and pragmatic attitudes in therapeutic trials. *Journal of Chronic Diseases* **20,** 737–648.

Smith, F. (1996). Mixed-model analysis of incomplete longitudinal data from a high-dose trial of Tacrine (Cognex®) in Alzheimer's patients. *Journal of Biopharmaceutical Statistics* **6,** 59–67.

Watkins, P. B., Zimmerman, H. J., Knapp, M. J., Gracon, S. I., and Lewis, K. W. (1994). Hepatotoxic effects of Tacrine administration in patients with Alzheimer's disease. *Journal of the American Medical Association* **271**(13), 992–998.

Wu, M. C. and Bailey, K. R. (1989). Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. *Biometrics* **45,** 939–955.

Wu, M. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44,** 175–188.

Received December 1995; *revised March* 1996; *accepted April* 1996.