STATISTICS IN MEDICINE
Statist. Med. 2001; **20**:2383–2397 (DOI: 10.1002/sim.904)

Group comparisons involving missing data in clinical trials: a comparison of estimates and power (size) for some simple approaches

Michael E. Miller[1,*,†], Timothy M. Morgan[1], Mark A. Espeland[1] and Scott S. Emerson[2]

[1]*Section on Biostatistics, Department of Public Health Sciences, Wake Forest University School of Medicine, Medical Center Blvd., Winston-Salem, NC 27157-1063, U.S.A.*
[2]*Department of Biostatistics, University of Washington, Seattle, WA 98195-0001, U.S.A.*

SUMMARY

When using 'intent-to-treat' approaches to compare outcomes between groups in clinical trials, analysts face a decision regarding how to account for missing observations. Most model-based approaches can be summarized as a process whereby the analyst makes assumptions about the distribution of the missing data in an attempt to obtain unbiased estimates that are based on functions of the observed data. Although pointed out by Rubin as often leading to biased estimates of variances, an alternative approach that continues to appear in the applied literature is to use fixed-value imputation of means for missing observations. The purpose of this paper is to provide illustrations of how several fixed-value mean imputation schemes can be formulated in terms of general linear models that characterize the means of distributions of missing observations in terms of the means of the distributions of observed data. We show that several fixed-value imputation strategies will result in estimated intervention effects that correspond to maximum likelihood estimates obtained under analogous assumptions. If the missing data process has been correctly characterized, hypothesis tests based on variances estimated using maximum likelihood techniques asymptotically have the correct size. In contrast, hypothesis tests performed using the uncorrected variance, obtained by applying standard complete data formula to singly imputed data, can provide either conservative or anticonservative results. Surprisingly, under several non-ignorable non-response scenarios, maximum likelihood based analyses can yield equivalent hypothesis tests to those obtained when analysing only the observed data. Copyright © 2001 John Wiley & Sons, Ltd.

1. INTRODUCTION

It has become traditional in clinical trials to adopt an 'intent-to-treat' approach for testing primary hypotheses in which all individuals are included in analyses according to the intervention group they were randomly assigned to, regardless of the completeness of their

*Correspondence to: Michael E. Miller, Wake Forest University School of Medicine, Department of Public Health Sciences, Medical Center Boulevard, Winston-Salem, NC 27157-1063, U.S.A.
†E-mail: mmiller@wfubmc.edu

Contract/grant sponsor: NIH; contract/grant number: N01HC45138, 5P60AG10484

Received August 1998
Accepted October 2000

follow-up or their adherence. The rationale for this approach is not based on explicit models or decision theory. Instead, the practice is justified as preserving the benefits of randomization and guarding against worst case scenarios, which are reasonable and defensible aims.

The intent-to-treat approach, since it requires the inclusion of participants who do not adhere (or who are suspected of not adhering), adds emphasis to the treatment of missing data since adherence and missing outcomes are often linked. Many proposals for accounting for missing outcomes when making intervention group comparisons of a single outcome from a controlled clinical trial have been made [1–4]. These include: (i) ignoring the missing data process and analysing only the observed data using standard complete case analyses; (ii) imputing values for the missing observations and using complete data analysis techniques [2, 3, 5–8], and (iii) postulating an underlying process that may be causing the observations to be missing, thus potentially necessitating the use of either selection or pattern-mixture models for analysing the data [4, 9]. To select an optimal statistical approach, the analyst must consider the underlying process that may have led to the missing observations. However, consideration of this process is often replaced by a vaguely justified strategy of imputing missing data in a manner chosen to be 'conservative' or to protect against 'worst case scenarios' [5–8, 10]. Such strategies may have good statistical properties only when missing data occur under narrowly circumscribed circumstances.

In this paper we examine separately the estimation of intervention effects and standard errors associated with strategies in which missing data are replaced with fixed, rather than randomly selected, values that are chosen based on intervention assignment and observed data. We first demonstrate that imputation strategies of this kind that have been proposed can be described by general linear models involving combinations of means of outcome measures. This allows model-based maximum likelihood to serve as a general basis for comparing the statistical efficiency and size of inference of alternative strategies.

We report several interesting findings. First, imputation-based strategies in which variances are not corrected for imputation of missing data can provide either conservative or anticonservative results, a somewhat counterintuitive finding since it is thought by many analysts that applying complete data analysis techniques to data containing imputations always produces variances that are too small. Second, several fixed-value imputation strategies will result in intervention effect estimates that correspond to maximum likelihood estimates under an analogous underlying model. Third, under several non-ignorable non-response scenarios, the strategy of including only individuals with complete data in analyses can yield equivalent hypothesis tests to those obtained using maximum likelihood based strategies. While the general methods that we present are not new, we are not aware of any publications that have explicitly drawn connections between results obtained from these commonly used fixed-value imputation techniques and the more appropriate maximum likelihood techniques.

2. IMPLICATIONS OF NON-ADHERENCE AND LOSS TO FOLLOW-UP

Analytical methods for making comparisons between groups containing missing observations vary with respect to estimation bias and efficiency. Although not always explicitly stated, most analytical approaches are based on an underlying model with associated assumptions. For instance, even in the simplest two-sample situation, the analyst can correctly ignore the missing data process only when the probability of a missing observation is independent of

the missed outcome [9]. Non-ignorable non-response occurs when the probability of response depends on the unobserved outcome. In this situation, assumptions regarding the missing data process, which often are not directly verifiable, typically are necessary to provide valid estimates and inference.

In many clinical trials, it is unlikely that the likelihood of a missing observation is unrelated to the assigned intervention and/or the unobserved outcome. We will refer to the situation where the likelihood of missing observations is unrelated to any observed or unobserved information as scenario 1. More often, the likelihood of a missing observation is linked to adherence, and thus to any intervention effects and possibly to unobserved outcomes [11]. In such situations, analysing only the observed data will often lead to inflated estimates of intervention effects. We consider three such scenarios, which we will describe for a two-armed trial of a control intervention A versus an active intervention B:

> *Scenario 2*. In the control intervention (A), which may be assumed to have no beneficial effect, lost follow-up occurs for participants who initiate some form of active therapy outside the study protocol and thus receive some partial benefits. In intervention B, assumed to be associated with a beneficial effect, lost follow-up occurs for participants who only partially comply with the treatment regimen and thus lose some potential benefits of therapy.
>
> *Scenario 3*. Participants in the control group who are lost accrue no benefit. Intervention group participants who are lost are those who cannot tolerate or are otherwise unresponsive to the active therapy (and also receive no benefit).
>
> *Scenario 4*. Participants in the control group who are lost to follow-up initiate and fully benefit from active therapy. Intervention group participants who are lost to follow-up are those who refuse this active therapy and thus receive no benefit from randomization to this study arm.

As stated by Rubin (reference [12], p. 155), 'without external information there will be no way to judge whether the non-respondents' missing values are systematically different from the respondents' observed values'. However, knowledge about adherence/patient characteristics and their relationships to missing outcomes can help to define strategies to obtain unbiased estimates of intervention differences from scenarios such as those described above. Under scenario 2, missing data might be considered to occur among individuals who have outcomes that are intermediate between those of control intervention A and active B. The values selected to impute for missing data would be chosen to be within the range between the mean outcomes observed for A and B. Imputation under scenario 3 might be devised to protect against bias that would result from non-response among the non-adherers in the active arm; thus, the average of observed outcomes from arm A might be imputed for missing observations in either arm. Scenario 4 may be described as a 'worst case' situation in which distributions of outcomes for lost participants are identical to those assigned to the opposite intervention. Here, the average of observed outcomes in each trial arm could be imputed for missing outcomes in the opposite arm. Similar schemes for single imputation of missing discrete data from clinical trials are described in Wittes *et al.* [3], and have been used in practice in either primary or secondary analyses of recent clinical trials [5–8].

Scenarios 2–4 represent situations where there exist systematic differences between the distribution of outcomes (Y) of participants with observed versus missing data. In Section 3.2 we identify explicit linear models for intervention effects in which the means of distributions

of missing data may be assumed to be linear combinations of the means underlying the observed data. The described Scenarios 1–4 are represented as special cases of this general class of models. Similar to Shih and Quan [4], we concentrate on a two-sample design with continuous observations made at one follow-up time. In contrast to Shih and Quan [4], who emphasize the joint analysis of the probability of drop-outs and the conditional mean of a continuous measurement for completers, we consider inference on the unconditional, hypothetical complete-data mean. Depending on the type of drop-out that is hypothesized, arguments can be made in favour of inference that is either conditional or unconditional on whether an outcome is observed [13–15]. For instance, in cases where loss to follow-up results from death and the hypothetical unobserved value for the deceased participant is not legitimate to consider, then the conditional approach can be more appropriate. However, in scenarios like we have described, where outcomes could be legitimately collected for non-responders, the unconditional approach can provide more valid inference.

3. NOTATION AND SIMPLE MODELS FOR MISSING DATA

3.1. Notation

Consider a clinical trial where n_{A} and n_{B} participants are randomized to interventions A and B, which are to be compared with respect to a single outcome after a period of follow-up. Assume that (i) initial intervention assignment is known for N participants ($i = 1, \ldots, N$) and is recorded in the $(N \times 1)$ vector X containing elements x_i, and (ii) Y contains elements y_i and represents the $(N \times 1)$ vector of outcomes, some of which may be missing. The total number of observations in each intervention group is denoted using n_{AO} and n_{BO}; the corresponding total number of missing observations is denoted using n_{AM} and n_{BM}. The $(N \times 1)$ vector M contains indicator variables denoting whether observations are missing ($M_{\mathrm{A}i} = 1$ and $M_{\mathrm{B}i} = 1$) or present ($M_{\mathrm{A}i} = 0$ and $M_{\mathrm{B}i} = 0$). Thus, $n_{\mathrm{AM}} = \sum_{j=1}^{n_{\mathrm{A}}} M_{\mathrm{A}j}$, $n_{\mathrm{BM}} = \sum_{j=1}^{n_{\mathrm{B}}} M_{\mathrm{B}j}$, $n_{\mathrm{A}} = n_{\mathrm{AO}} + n_{\mathrm{AM}}$, $n_{\mathrm{B}} = n_{\mathrm{BO}} + n_{\mathrm{BM}}$, and $N = n_{\mathrm{A}} + n_{\mathrm{B}}$. Lastly, allow $\bar{Y}_{\mathrm{AO}}$ to represent the observed mean for intervention A, $\bar{Y}_{\mathrm{BO}}$ the observed mean for intervention B, $\hat{\delta} = \bar{Y}_{\mathrm{AO}} - \bar{Y}_{\mathrm{BO}}$ the observed difference, and $\hat{\mathrm{var}}(\hat{\delta}) = s^2_{\mathrm{obs}}(1/n_{\mathrm{AO}} + 1/n_{\mathrm{BO}})$, where s^2_{obs} is the pooled variance based on the observed data in groups A and B.

3.2. A simple model for characterizing missing data

Many analytical methods for addressing missing outcomes correspond to characterizing parameters of the distributions of the missing outcomes as a function of the parameters of the distributions of the observed outcomes. Underlying beliefs regarding the nature of the missing data process or conservative philosophies [3] can be represented by expressing the means of the missing data distributions as linear combinations of the means of the observed data distributions. Let μ_{AO} and μ_{BO} denote means for the distribution of observed data from the two study arms and define

$$\begin{aligned} \mu_{\mathrm{AM}} &= \alpha_{\mathrm{A}} + \beta_{\mathrm{A}}\mu_{\mathrm{AO}} + \gamma_{\mathrm{A}}\mu_{\mathrm{BO}} \\ \mu_{\mathrm{BM}} &= \alpha_{\mathrm{B}} + \beta_{\mathrm{B}}\mu_{\mathrm{AO}} + \gamma_{\mathrm{B}}\mu_{\mathrm{BO}} \end{aligned} \tag{1}$$

The coefficients of these expressions can be selected to correspond to a range of possible assumptions regarding the missing data process, including scenarios 1–4, above. For example, if one suspects scenario 1, then the underlying assumptions that $\mu_{AM} = \mu_{AO}$ and $\mu_{BM} = \mu_{BO}$ would be represented by $\alpha_A = 0,\ \beta_A = 1,\ \gamma_A = 0,\ \alpha_B = 0,\ \beta_B = 0,\ \gamma_B = 1$. If scenario 2 is suspected, then the assumption that $\mu_{AM} = \mu_{BM} = (\mu_{AO} + \mu_{BO})/2$ would be represented by $\alpha_A = 0,\ \beta_A = 0.5,\ \gamma_A = 0.5,\ \alpha_B = 0,\ \beta_B = 0.5,\ \gamma_B = 0.5$. Likewise, for scenario 3, the assumption that $\mu_{AM} = \mu_{BM} = \mu_{AO}$ corresponds to $\alpha_A = 0,\ \beta_A = 1,\ \gamma_A = 0,\ \alpha_B = 0,\ \beta_B = 1,\ \gamma_B = 0$. Finally, if scenario 4 is adopted, then $\mu_{AM} = \mu_{BO},\ \mu_{BM} = \mu_{AO}$, and $\alpha_A = 0,\ \beta_A = 0,\ \gamma_A = 1,\ \alpha_B = 0,\ \beta_B = 0,\ \gamma_B = 1$. For clinical trials, careful thought and clinical input is vital for appropriately characterizing the distributions of missing outcomes. If correct assumptions about the missing data process are made, then unbiased estimates of intervention differences are possible.

4. ESTIMATION OF INTERVENTION DIFFERENCES

We contrast two approaches for developing estimates for the means of the missing data distributions discussed above: maximum likelihood (ML) and fixed-value imputation (FVI). Our intent is to demonstrate the concordance of estimates of the relative intervention effect from these two approaches based on models such as (1).

4.1. Maximum likelihood estimation

Shih and Quan [4] detail the theory of maximum likelihood estimation using a pattern-mixture model approach to factor the likelihood for incomplete data obtained from a clinical trial. If (φ, π) are unknown parameters and X is fixed, the joint probability of Y and M can be expressed as

$$P(Y, M \mid X, \varphi, \pi) = P(M \mid X, \pi) P(Y \mid M, X, \varphi)$$

Missing data patterns need to be specified in order to use the pattern-mixture approach. For our simple two-sample situation, we can form four missing data patterns (that is, two patterns representing missing or non-missing data within intervention groups A and B, respectively). We index these patterns using $k = 1, 2, 3, 4$, for intervention A-observed data (AO), intervention A-missing data (AM), intervention B-observed data (BO) and intervention B-missing data (BM), respectively. For convenience, at times we use the numerical subscripts 1–4 in place of the coded subscripts AO, AM, BO and BM. Assume that for group A, $M_{Ai} \sim_{\text{ind}} \text{Bernoulli}(\pi_{AM})$, and for group B, $M_{Bi} \sim_{\text{ind}} \text{Bernoulli}(\pi_{BM})$. If $P(Y \mid M, X, \varphi) \sim_{\text{ind}} \text{Normal}(\varphi_{(k)})$, $\varphi = (\varphi_{AO}, \varphi_{AM}, \varphi_{BO}, \varphi_{BM})$, and $\varphi_{(k)} = (\mu_{(k)}, \sigma^2_{(k)})$, then the likelihood for the identified parameters has the form

$$L(\pi_{AO}, \pi_{AM}, \pi_{BO}, \pi_{BM}, \varphi_{AO}, \varphi_{BO}) = \pi_{AO}^{n_{AO}} \pi_{AM}^{n_{AM}} \pi_{BO}^{n_{BO}} \pi_{BM}^{n_{BM}} \prod_{j=1}^{n_{AO}} N(y_j \mid \mu_{AO}, \sigma^2_{AO}) \times \prod_{j=1}^{n_{BO}} N(y_j \mid \mu_{BO}, \sigma^2_{BO}) \quad (2)$$

where $\pi_{AO} = 1 - \pi_{AM}$ and $\pi_{BO} = 1 - \pi_{BM}$. As Little [16] indicates, when the parameters in φ and π are distinct and the log-likelihood can be factored into separate components each containing unique parameters, then non-iterative maximum likelihood (ML) estimates can often be obtained. For likelihood (2), ML estimates of the true means for interventions A and B are $\hat{\mu}_A = \hat{\pi}_{AO}\hat{\mu}_{AO} + \hat{\pi}_{AM}\hat{\mu}_{AM}$ and $\hat{\mu}_B = \hat{\pi}_{BO}\hat{\mu}_{BO} + \hat{\pi}_{BM}\hat{\mu}_{BM}$. The ML estimate of the intervention difference is $\hat{\Delta} = \hat{\mu}_A - \hat{\mu}_B$.

Because $(\mu_{AM}, \sigma^2_{AM})$ and $(\mu_{BM}, \sigma^2_{BM})$ in $P(Y \mid M, X, \varphi)$ are not identified, a unique solution to the above equations cannot be obtained without further assumptions to reduce the number of parameters. These assumptions are referred to as identifying restrictions by Little [16]. Different sets of identifying restrictions give different estimates of the intervention effects. The general linear model specified in (1) represents one functional relationship between the means of the missing and observed distributions. This model can be used to specify the identifying restrictions for the means of the distributions of missing data. Many different missing data processes can be represented through the choice of the underlying coefficients in (1). By combining the probabilities of 'missingness' with the relationships defined by the parameters in equation (1), the following expressions for the average outcomes for populations receiving interventions A and B are obtained:

$$\begin{aligned} \mu_A &= \pi_{AM}\alpha_A + (1 - \pi_{AM} + \pi_{AM}\beta_A)\mu_{AO} + \pi_{AM}\gamma_A\mu_{BO} \\ \mu_B &= \pi_{BM}\alpha_B + \pi_{BM}\beta_B\mu_{AO} + (1 - \pi_{BM} + \pi_{BM}\gamma_B)\mu_{BO} \end{aligned} \tag{3}$$

The overall intervention effect can be written as

$$\begin{aligned} \mu_A - \mu_B = (\pi_{AM}\alpha_A - \pi_{BM}\alpha_B) + (1 - \pi_{AM} + \pi_{AM}\beta_A - \pi_{BM}\beta_B)\mu_{AO} \\ - (1 - \pi_{BM} + \pi_{BM}\gamma_B - \pi_{AM}\gamma_A)\mu_{BO} \end{aligned} \tag{4}$$

Maximum likelihood estimates of this effect can be obtained by substituting the MLEs for μ_{AO}, μ_{BO}, π_{AM} and π_{BM}. The MLEs for μ_{AO} and μ_{BO} are $\bar{Y}_{AO}$ and $\bar{Y}_{BO}$, whereas the MLEs of π_{AM} and π_{BM} are $n_{AM}/(n_{AO} + n_{AM})$ and $n_{BM}/(n_{BO} + n_{BM})$.

4.2. Fixed-value imputation

In fixed-value imputation, each missing data point is replaced according to a selected strategy and the intervention effect is estimated as the difference between the means of the augmented data (observed and imputed) in each intervention. For scenarios 1–4, the associated fixed-imputation procedure corresponds to the following substitution rules:

1. Observed means from interventions A and B are substituted for missing outcomes in the arm that the participant was assigned (strategy 1).
2. Average of the observed means from interventions A and B is substituted for missing outcomes in both intervention arms (strategy 2).
3. Observed mean from arm A (the control arm) is substituted for missing outcomes in both intervention arms (strategy 3).
4. Mean in each intervention group is substituted for missing data in the other group (strategy 4).

Each of the mean values imputed in strategies 1–4 can also be represented using equation (1) and the parameters used to express the corresponding scenarios 1–4 in Section 3.2. Importantly, the application of complete data formula to the data containing fixed-value imputations using strategies 1–4 results in moment estimates for intervention effects that are exactly the same as the maximum likelihood estimates obtained from equation (4). This is easily demonstrated by expressing the intervention means obtained using complete data formula as weighted linear combinations of the means for the observed and imputed data. Fixed-value imputation estimates of the means can be expressed as

$$\hat{\mu}_{\mathrm{A}}^{\mathrm{FVI}} = \left(\frac{n_{\mathrm{AO}}}{n_{\mathrm{AO}} + n_{\mathrm{AM}}}\right)\bar{Y}_{\mathrm{AO}} + \left(\frac{n_{\mathrm{AM}}}{n_{\mathrm{AO}} + n_{\mathrm{AM}}}\right)(\alpha_{\mathrm{A}} + \beta_{\mathrm{A}}\bar{Y}_{\mathrm{AO}} + \gamma_{\mathrm{A}}\bar{Y}_{\mathrm{BO}})$$

$$\hat{\mu}_{\mathrm{B}}^{\mathrm{FVI}} = \left(\frac{n_{\mathrm{BO}}}{n_{\mathrm{BO}} + n_{\mathrm{BM}}}\right)\bar{Y}_{\mathrm{BO}} + \left(\frac{n_{\mathrm{BM}}}{n_{\mathrm{BO}} + n_{\mathrm{BM}}}\right)(\alpha_{\mathrm{B}} + \beta_{\mathrm{B}}\bar{Y}_{\mathrm{AO}} + \gamma_{\mathrm{B}}\bar{Y}_{\mathrm{BO}})$$

By calculating $\hat{\Delta}^{\mathrm{FVI}} = \hat{\mu}_{\mathrm{A}}^{\mathrm{FVI}} - \hat{\mu}_{\mathrm{B}}^{\mathrm{FVI}}$ and applying a little algebra, the MLE estimate of $\Delta = \mu_{\mathrm{A}} - \mu_{\mathrm{B}}$ as defined by equation (4) is obtained.

5. VARIANCE CALCULATIONS

When FVI strategies, rather than models, are used to account for missing observations, a basis is not available for developing appropriate standard errors of estimated intervention differences. A not uncommon strategy is to adopt fixed-value imputation based on a particular strategy and calculate variances and standard errors as if the imputed data were actually observed. Depending on the imputation strategy chosen, this practice can lead to variances estimates that may be either too large or too small. In contrast, model-based likelihood methods provide a pathway for obtaining variances that asymptotically result in confidence intervals with the chosen coverage probability.

Formulae for obtaining estimates of the variances of maximum likelihood estimates of the intervention effect can be obtained using multivariate Taylor series expansions. These variance estimates contain terms related to the variance of both the estimates of the proportion missing (binomial variation) and the variability of the observed measurements in each intervention. Using model (1) to define restrictions on means, and assuming $\sigma_{\mathrm{AM}}^2 = \sigma_{\mathrm{AO}}^2$, $\sigma_{\mathrm{BM}}^2 = \sigma_{\mathrm{BO}}^2$, the estimated variance of the intervention effect takes the form

$$V_{\mathrm{ML}}(\hat{\Delta}) = K_1^2\frac{\hat{\sigma}_{\mathrm{AO}}^2}{n_{\mathrm{AO}}} + K_2^2\frac{\hat{\sigma}_{\mathrm{BO}}^2}{n_{\mathrm{BO}}} + K_3^2\frac{\hat{\pi}_{\mathrm{AO}}(1-\hat{\pi}_{\mathrm{AO}})}{n_{\mathrm{A}}} + K_4^2\frac{\hat{\pi}_{\mathrm{BO}}(1-\hat{\pi}_{\mathrm{BO}})}{n_{\mathrm{B}}} \tag{5}$$

where

$$K_1 = (1 - \hat{\pi}_{\mathrm{AM}} + \hat{\pi}_{\mathrm{AM}}\beta_{\mathrm{A}} - \hat{\pi}_{\mathrm{BM}}\beta_{\mathrm{B}})$$

$$K_2 = (1 - \hat{\pi}_{\mathrm{BM}} + \hat{\pi}_{\mathrm{BM}}\gamma_{\mathrm{B}} - \hat{\pi}_{\mathrm{AM}}\gamma_{\mathrm{A}})$$

$$K_3 = (\alpha_{\mathrm{A}} + (\beta_{\mathrm{A}} - 1)\hat{\mu}_{\mathrm{AO}} + \gamma_{\mathrm{A}}\hat{\mu}_{\mathrm{BO}})$$

$$K_4 = (-\alpha_{\mathrm{B}} - \beta_{\mathrm{B}}\hat{\mu}_{\mathrm{AO}} + (1 - \gamma_{\mathrm{B}})\hat{\mu}_{\mathrm{BO}})$$

In the remainder of this section, we contrast this expression for the estimated variance with the naive, uncorrected estimate of variance obtained using FVI, expressed as

$$V_{\mathrm{FVI}}(\hat{\Delta}) = \left(\frac{n_{\mathrm{A}} + n_{\mathrm{B}}}{n_{\mathrm{A}} n_{\mathrm{B}}}\right)\left(\frac{1}{n_{\mathrm{A}} + n_{\mathrm{B}} - 2}\right)((n_{\mathrm{AO}} + n_{\mathrm{BO}} - 2)s^2_{\mathrm{obs}} + n_{\mathrm{AO}}\hat{\pi}_{\mathrm{AM}}K_3^2 + n_{\mathrm{BO}}\hat{\pi}_{\mathrm{BM}}K_4^2) \quad (6)$$

We do so not to advocate in any way adopting uncorrected variances, but to indicate that this approach can have surprisingly diverse consequences for relative size of inference, sometimes being extremely conservative and other times being quite liberal. This variance estimator can be expressed as a linear function of s^2_{obs} and $\hat{\delta}^2$ for each of the four strategies we have considered.

Table I summarizes the models underlying the four strategies and provides in column 4 the estimator of the intervention effect (which, as noted in Section 4.2, is equivalent for FVI and ML estimation). Each of these estimators has been expressed as a function of the observed intervention effect. The assumptions corresponding to the model specified in equation (1) and the associated restrictions on the pattern-mixture model are listed in columns 2 and 3 of Table I.

The ratio of the ML variance estimate (V_{ML}) to the naive, uncorrected variance estimate (V_{FVI}) can be approximated by

$$\frac{V_{\mathrm{ML}}}{V_{\mathrm{FVI}}} = \frac{(\frac{K_1^2}{\hat{\pi}_{\mathrm{AO}}} + \frac{K_2^2}{\hat{\pi}_{\mathrm{BO}}})s^2_{\mathrm{obs}} + \hat{\pi}_{\mathrm{AO}}\hat{\pi}_{\mathrm{AM}}K_3^2 + \hat{\pi}_{\mathrm{BO}}\hat{\pi}_{\mathrm{BM}}K_4^2}{(\hat{\pi}_{\mathrm{AO}} + \hat{\pi}_{\mathrm{BO}})s^2_{\mathrm{obs}} + \hat{\pi}_{\mathrm{AO}}\hat{\pi}_{\mathrm{AM}}K_3^2 + \hat{\pi}_{\mathrm{BO}}\hat{\pi}_{\mathrm{BM}}K_4^2} \quad (7)$$

where we assume $n_{\mathrm{A}} = n_{\mathrm{B}}$, $\hat{\sigma}^2 = \hat{\sigma}^2_{\mathrm{A}} = \hat{\sigma}^2_{\mathrm{B}}$, $s^2_{\mathrm{obs}} \cong \hat{\sigma}^2$, and the quantities $n_{\mathrm{AO}} - 1$, $n_{\mathrm{BO}} - 1$, $n_{\mathrm{A}} - 1$ and $n_{\mathrm{B}} - 1$ are approximated by n_{AO}, n_{BO}, n_{A} and n_{B}. Relative to maximum likelihood, the use of V_{FVI} will yield the nominal probabilities of type I error when the ratio of the variance estimates is 1, will be inappropriately anticonservative (liberal) when $\hat{\pi}_{\mathrm{BO}}K_1^2 + \hat{\pi}_{\mathrm{AO}}K_2^2 > \hat{\pi}_{\mathrm{AO}}\hat{\pi}_{\mathrm{BO}}(\hat{\pi}_{\mathrm{AO}} + \hat{\pi}_{\mathrm{BO}})$, and will be conservative when $\hat{\pi}_{\mathrm{BO}}K_1^2 + \hat{\pi}_{\mathrm{AO}}K_2^2 < \hat{\pi}_{\mathrm{AO}}\hat{\pi}_{\mathrm{BO}}(\hat{\pi}_{\mathrm{AO}} + \hat{\pi}_{\mathrm{BO}})$. Thus, the operating properties of hypothesis tests using FVI are governed by the underlying proportion of data observed in each intervention group and the assumptions that are made relating the means of the distributions of the unobserved data to the corresponding means of the observed data.

For each of the scenarios we consider, Table I provides in column 5 this ratio for the special case of $\hat{\pi}_{\mathrm{O}} = \hat{\pi}_{\mathrm{AO}} = \hat{\pi}_{\mathrm{BO}}$, for which expressions simplify. In this column, R represents the ratio of the observed intervention effect to the observed pooled standard deviation. As intuition might lead us to believe, use of V_{FVI} under imputation strategy 1 produces hypothesis tests that are anticonservative, since in the presence of missing data, V_{FVI} will always be smaller than V_{ML}. When $\hat{\pi}_{\mathrm{AO}} \neq \hat{\pi}_{\mathrm{BO}} \neq 1$, then use of V_{FVI} under strategy 1 will be anticonservative since $\hat{\pi}_{\mathrm{AO}}\hat{\pi}_{\mathrm{BO}} < 1$. For $\hat{\pi}_{\mathrm{AO}} = \hat{\pi}_{\mathrm{BO}}$, use of V_{FVI} under imputation strategies 2 and 3 results in $V_{\mathrm{ML}} = V_{\mathrm{FVI}}$ and a ratio of 1. For $\hat{\pi}_{\mathrm{AO}} \neq \hat{\pi}_{\mathrm{BO}}$, use of V_{FVI} will always be anticonservative for strategy 2; whereas, for strategy 3 the test using V_{FVI} is anticonservative when $\hat{\pi}_{\mathrm{AO}} < \hat{\pi}_{\mathrm{BO}}$ and conservative when $\hat{\pi}_{\mathrm{AO}} > \hat{\pi}_{\mathrm{BO}}$. For strategy 4, the sign of the estimated intervention effect does not change as long as $\hat{\pi}_{\mathrm{AO}} + \hat{\pi}_{\mathrm{BO}} > 1$. In this situation, use of V_{FVI} will always be conservative. We would be reluctant to use any imputation strategy if the proportion of observed data is small in either group. In Table II, we provide a summary of how hypothesis tests using FVI perform in comparison to MLE under each strategy.

Table I. Estimators of intervention differences and ratios of variance under four imputation strategies.

Missing data strategy*	Missing data model assumptions	Pattern-mixture restrictions	Estimator of intervention difference	$\frac{V_{ML}}{V_{FVI}}$ if $\hat{\pi}_O = \hat{\pi}_{AO} = \hat{\pi}_{BO}$ $(R = \hat{\delta}/s_{obs})$
1	$\alpha_A = \alpha_B = 0,$ $\beta_A = 1,\ \beta_B = 0,$ $\gamma_A = 0,\ \gamma_B = 1$	$\mu_{AM} = \mu_{AO}, \mu_{BM} = \mu_{BO},$ $\sigma^2_{AM} = \sigma^2_{AO},\ \sigma^2_{BM} = \sigma^2_{BO}$	$\hat{\delta}$	$\frac{1}{\hat{\pi}^2_O}$
2	$\alpha_A = \alpha_B = 0,$ $\beta_A = \beta_B = 0.5,$ $\gamma_A = \gamma_B = 0.5$	$\mu_{AM} = \mu_{BM} = 0.5(\mu_{BO} + \mu_{AO}),$ $\sigma^2_{AM} = \sigma^2_{AO},\ \sigma^2_{BM} = \sigma^2_{BO}$	$\left(\frac{\hat{\pi}_{AO} + \hat{\pi}_{BO}}{2}\right)\hat{\delta}$	1
3	$\alpha_A = \alpha_B = 0,$ $\beta_A = \beta_B = 1,$ $\gamma_A = \gamma_B = 0$	$\mu_{AM} = \mu_{BM} = \mu_{AO},$ $\sigma^2_{AM} = \sigma^2_{AO},\ \sigma^2_{BM} = \sigma^2_{BO}$	$\hat{\pi}_{BO}\hat{\delta}$	1
4	$\alpha_A = \alpha_B = 0,$ $\beta_A = 0,\ \beta_B = 1,$ $\gamma_A = 1,\ \gamma_B = 0$	$\mu_{AM} = \mu_{BO},\ \mu_{BM} = \mu_{AO},$ $\sigma^2_{AM} = \sigma^2_{AO},\ \sigma^2_{BM} = \sigma^2_{BO}$	$(\hat{\pi}_{AO} + \hat{\pi}_{BO} - 1)\hat{\delta}$	$\frac{(2\hat{\pi}_O - 1)^2 + \hat{\pi}^2_O(1 - \hat{\pi}_O)R^2}{\hat{\pi}^2_O + \hat{\pi}^2_O(1 - \hat{\pi}_O)R^2}$

* 1. Observed means from interventions A and B for missing responses in the arm that the participant was assigned.
2. Average of observed means from interventions A and B for missing responses in both intervention arms.
3. The observed mean from intervention A (potentially a control arm) for missing responses in both intervention arms.
4. The observed means from interventions A and B for missing responses in the opposite arm (a cross-over scheme).

Table II. Summary of performance of hypothesis tests using V_{FVI} versus V_{ML}.

Imputation strategy*	Performance of V_{FVI} versus V_{ML}		
	Conservative	Correct	Anticonservative
1	Never	Never	Always
2	Never	$\hat{\pi}_{AO} = \hat{\pi}_{BO}$	$\hat{\pi}_{AO} \neq \hat{\pi}_{BO}$
3	$\hat{\pi}_{AO} > \hat{\pi}_{BO}$	$\hat{\pi}_{AO} = \hat{\pi}_{BO}$	$\hat{\pi}_{AO} < \hat{\pi}_{BO}$
4 when $\hat{\pi}_{AO} + \hat{\pi}_{BO} > 1$	Always	Never	Never

* 1. Observed means from interventions A and B for missing responses in the arm that the participant was assigned.
2. Average of observed means from interventions A and B for missing responses in both intervention arms.
3. The observed mean from intervention A (potentially a control arm) for missing responses in both intervention arms.
4. The observed means from interventions A and B for missing responses in the opposite arm (a cross-over scheme).

Hypothesis tests based on V_{ML} can result, at times, in the same significance levels as obtained by analysing only the observed data for each of the four mean imputation schemes considered. The reason for this can be determined by inspection of equation (4). When α_A and α_B both equal zero and the coefficient associated with μ_{AO} (K_1 from above) is equal to the coefficient associated with μ_{BO} (K_2 from above), then the estimated complete data intervention effect will be directly proportional to the observed intervention effect (that is, $K\hat{\delta}$ where $K = K_1 = K_2$). Thus, K represents the amount of 'shrinkage' of the observed intervention effect in comparison to the estimated complete data effect. When the estimate of the intervention effect can be expressed as $K\hat{\delta}$, V_{ML} may also reduce to the square of the proportionality constant times the observed variance $K^2\,\hat{\text{var}}(\hat{\delta})$, thus reducing the hypothesis test to a test based solely on the observed data. This relationship is easy to verify through inspection of equation (5) when $n = n_A = n_B$ for imputation strategy 1, since $K = K_1 = K_2 = 1$ and $K_3 = K_4 = 0$. For strategies 2–3, the relationship also holds when $n = n_A = n_B$ and $(\hat{\delta}^2/n) \to 0$. In contrast, for strategy 4 it also is necessary to restrict attention to situations where $\hat{\pi}_{AO} + \hat{\pi}_{BO} > 1$. In summary, for imputation strategies 1–3, an intervention difference exists in the observed data if and only if an intervention difference also exists in the full data (that is, for all subjects irrespective of missingness). For strategy 4, this statement holds when $\hat{\pi}_{AO} + \hat{\pi}_{BO} > 1$.

6. SIMULATION RESULTS

To investigate the consequences of use of V_{FVI} versus V_{ML} on the size and power of the test for an intervention effect, we assumed $n_A = n_B = 100$ and performed a small simulation study. Data were generated from a normal distribution with unit variance assuming that each of the four models corresponding to the missing data scenarios previously described were true. Assuming that $\Delta = \mu_A - \mu_B$ was the true underlying intervention effect in the observed data, hypothesis tests for H_0: $\mu_A - \mu_B = 0$ were carried out using two-sided level 0.05 tests. To represent a realistic situation that might be encountered in a clinical trial, we assumed $\pi_{AM} = \pi_{BM} = 0.2$ and allowed Δ to take the values 0, 0.1, 0.2 and 0.3. Mean estimated intervention differences

Table III. Results from 10 000 simulations to contrast analytical approaches for handling missing observations.

Missing data scenario and imputation strategy*	Δ	Estimated intervention difference		Power (size)			V_{ML}/V_{FVI}
		Observed data	Using FVI or ML	Observed data	Using V_{FVI}	Using V_{ML}	
1	0.0	−0.0003	−0.0003	0.0539	0.1314	0.0558	1.5562
	0.1	0.1037	0.1037	0.0998	0.1913	0.1025	1.5537
	0.2	0.1971	0.1971	0.2395	0.3754	0.2419	1.5529
	0.3	0.2997	0.2997	0.4708	0.6264	0.4764	1.5531
2	0.0	0.0010	0.0008	0.0532	0.0530	0.0542	0.9913
	0.1	0.1238	0.0991	0.1226	0.1222	0.1240	0.9912
	0.2	0.2510	0.2009	0.3584	0.3586	0.3626	0.9912
	0.3	0.3752	0.3002	0.6632	0.6629	0.6666	0.9912
3	0.0	−0.0008	−0.0008	0.0534	0.0527	0.0536	0.9924
	0.1	0.1275	0.1021	0.1287	0.1276	0.1289	0.9935
	0.2	0.2504	0.2003	0.3463	0.3432	0.3477	0.9917
	0.3	0.3759	0.3006	0.6585	0.6566	0.6598	0.9936
4	0.0	0.0005	0.0004	0.0530	0.0100	0.0500	0.5596
	0.1	0.1680	0.1007	0.1892	0.0584	0.1828	0.5618
	0.2	0.3351	0.2015	0.5639	0.2968	0.5546	0.5702
	0.3	0.4996	0.2999	0.8780	0.6736	0.8743	0.5810

* For each scenario used to generate data, the corresponding imputation strategies were:
1. Observed means from intervention groups A and B for missing responses in group that the participant was assigned.
2. Average of observed means for missing data in both groups A and B.
3. Group A mean for missing data in both groups A and B.
4. Group A mean for missing data in group B and group B mean for missing data in group A.

and power (size) of the tests based on 10 000 simulations are presented in Table III for three hypothesis testing procedures that used: (i) only observed data, (ii) V_{FVI} obtained from singly imputed data from the imputation strategy corresponding to the underlying missing data scenario, and (iii) V_{ML} obtained from the pattern-mixture model corresponding to the underlying missing data scenario. This table also presents the average ratio of V_{ML} to V_{FVI} obtained from the 10 000 simulations.

The results in Table III clearly illustrate that the difference between the means of the observed outcomes for each intervention unbiasedly estimates the true expected differences between the interventions' outcomes only under scenario 1. FVI and ML provide the same unbiased estimate of the intervention difference in all situations considered. Under the null hypothesis of $\Delta = 0$, ML always gives the right type I error, while FVI does not under strategies 1 and 4. For strategy 4, use of the uncorrected variance can result in substantial loss of power for $\Delta > 0$ and inappropriate size for $\Delta = 0$. For strategy 4, the actual probability of type I error associated with use of V_{FVI} is 0.01, even though the analyst using this procedure would be reporting that the tests were carried out at the 0.05 level. As expected, for the scenarios considered, ML outperforms FVI for hypothesis testing while providing the same intervention effects that would be obtained using FVI. Finally, as shown algebraically in

Section 5, the ML technique and analyses of just the observed data provide equivalent power (size) for strategies 1–4, even though the estimated intervention effect can vary greatly using these two approaches.

7. EXAMPLE

The Fitness, Arthritis and Seniors Trial (FAST) was a clinical trial with the aim of determining the effects of selected exercise programmes on pain and disability in older adults with knee osteoarthritis [17]. Patients were randomized to one of three interventions at two clinical centres and were followed for up to 18 months. The three interventions that were used included an aerobics exercise programme, a resistance exercise programme and a health education programme. Primary outcome measures were collected at 3, 9 and 18 months follow-up and included performance measures of physical function and self-reported disability.

For purposes of illustrating the missing data techniques previously described, we only use data collected for 293 participants randomized to the health education ($N = 149$) and aerobics ($N = 144$) exercise programmes. The two outcome measures that we consider are: (i) the distance walked in six minutes and (ii) the time required to get in and out of a simulated car. The car outcome was collected only at a single clinical site, thus providing a smaller total sample size ($N = 138$) than that available for the distance walked measure. For illustrative purposes, we present comparisons between education and aerobics arms of the mean distance walked and the mean time to get in and out of the car at the 9 month follow-up visit.

Concentrating initially on the distance walked outcome, 23 per cent and 26 per cent of outcomes were missing at 9 months for the education and aerobics arms, respectively. The mean distance walked for patients observed at 9 months and receiving the health education intervention was 1361 feet (approx. 40.8 m), whereas the mean distance for those patients observed at 9 months and receiving the aerobics programe was 1468 feet (approx. 44.0 m). In Table IV, we present analyses of the intervention effect under five assumptions about the missing data process and using FVI and ML procedures.

For the distance walked outcome, the estimated intervention effect is shrunk toward zero from the observed -107.19 for imputation strategies 2–4. Strategy 4, which represents the most conservative assumptions considered regarding the missing data process, provides a 50 per cent reduction in the observed intervention effect. The t-statistics associated with V_{FVI} are almost identical for strategies 2 and 3. Based on the ratio presented in the last column of Table I, this result was predictable since the proportion of missing observations in each arm is almost identical. Ignoring the imputations when calculating variances using imputation strategy 4 produces a t-statistic of -1.70 ($p = 0.09$), resulting in no rejection of the null hypothesis under this strategy. In contrast, the ML procedure produces t-statistics that are relatively constant regardless of the underlying model. This result is not surprising. Based on the characteristics of the data for this example and the results discussed in the previous section, the ML hypothesis tests would be predicted to reduce to tests based on analysing only observed data. For strategies 1–4, the precision of the estimated intervention effects and the variance of these effects are almost completely dependent on the variance of the intervention effect estimated from the observed data and the multiplier used to shrink the observed estimate.

For the time to get in and out of the car outcome, 28 per cent and 34 per cent of outcomes were missing at 9 months in the education and aerobics arms, respectively. The mean time

Table IV. Results obtained applying missing data procedures to FAST data.

Variable	Imputation strategy*	Estimated intervention effect	Ignore imputations for variances		ML estimation	
			Variance	t-statistic	Variance	t-statistic
Distance walked	None	−107.19	1718.12	−2.59		
in 6 min at	1	−107.19	965.43	−3.45	1719.00	−2.59
9-month visit	2	−80.46	972.81	−2.58	975.84	−2.58
	3	−78.90	980.52	−2.52	946.96	−2.56
	4	−53.73	994.96	−1.70	461.19	−2.50
Time to get in	None	2.52	0.695	3.02		
and out of car	1	2.52	0.327	4.40	0.654	3.11
at 9-month visit	2	1.73	0.337	2.99	0.320	3.07
	3	1.70	0.347	2.89	0.320	3.01
	4	0.95	0.367	1.57	0.132	2.61

* 1. Observed means from education and aerobics arms for missing responses in arm that the participant was assigned.
2. Average of observed means for missing data in both education and aerobics intervention arms.
3. Education mean for missing data in both aerobics and education arms.
4. Education mean for missing data in aerobics arm and aerobics mean for missing data in education arm.

for the 49 observed outcomes on the education arm was 10.81 seconds, whereas the mean time for the 46 observed outcomes on the aerobics arm was 8.29 seconds. Results similar to those presented for the distance walked outcome are presented for this outcome in the bottom panel of Table IV. The estimated intervention effect under strategy 4 is approximately 40 per cent of the observed effect. In contrast to the relatively constant t-statistics obtained from the ML procedures under all four models, the p-values associated with the t-statistics obtained ignoring imputations when calculating variances range from <0.001 for strategy 1 to 0.120 for strategy 4.

Discussions with interventionists in the FAST study indicate that non-responders in the aerobics intervention were most likely to have not been participating in exercise and thus would have had outcomes similar to those observed for participants in the education intervention. Thus, the estimated intervention effect based on imputation strategy 3 may be the least biased. It is also believed that non-responders in the education intervention did not begin an exercise programme. Thus, in this situation, the cross-over imputation strategy produces an estimate that is most certainly too conservative.

8. DISCUSSION

How an analyst handles non-response during the analysis of clinical trials data can affect the overall conclusions derived from a study. When the probability of response depends on the unobserved outcome, the missing data process is non-ignorable and using the observed data in standard complete case analyses can lead to biased estimates. Unbiased estimates of intervention differences are possible if correct assumptions regarding the missing data process are made and these assumptions are used to form estimation functions of the observed data.

The choice of a specific imputation approach from the many possible strategies for imputation of missing observations in clinical trials is inherently based on some underlying assumptions regarding a model for the missing data process. Even if an analyst selects a specific approach with the primary intent of producing a conservative analysis, then a decision must be made regarding the degree of conservatism the analyst desires. In contrast to the four strategies we have considered, more realistic conservatism may be obtained by using non-zero values for α_A and α_B in model (1) to allow for the possibility that non-compliers with missing observations have outcomes that are generally worse than those observed in either intervention. However, without additional external information, there is little way of knowing what values α_A and α_B should take. Ultimately, this choice regarding the desired degree of conservatism also involves the presence of an underlying belief regarding a model for the missing data process.

We have illustrated how, under a class of models for the means of missing data, pattern-mixture models can produce estimates of intervention differences in the presence of missing outcomes that are identical to those obtained under various imputation procedures. However, the variances obtained from ML estimation using a pattern-mixture model and fixed-value imputation procedures are not identical. The latter procedure provides variance estimates of estimated intervention differences that can be quite biased.

For strategies 2, 3 and 4, the estimated intervention effects from either pattern-mixture models or imputation procedures 'penalize' a trial containing missing data by shrinking toward zero the estimated intervention difference that was observed. For each, the 'penalized' intervention effect was shown to be a multiple of the observed intervention effect. Thus, the overall impact of the intervention in terms of public health must be evaluated based on a diminished intervention effect. The precision of the 'penalized' estimate of the intervention effect is also dependent on the precision of the observed intervention effect. Use of biased variance estimates obtained by assuming the imputed observations are observed can ultimately result in confidence intervals with inaccurate coverage and hypothesis tests with size less than the nominally chosen level. An interesting finding in this paper is that for large samples, the hypothesis tests obtained from the ML approach can, under certain assumptions, reduce to tests that are identical to the test obtained if one were to analyse only the observed data.

In the absence of outcome data, the only way to make inference regarding the entire population available for randomization is to make assumptions regarding the distribution of missing data or the missing data process. In the face of uncertainty about this distribution, realistic assumptions should be made in order to provide estimates (sometimes conservative) that take into account the assumptions made regarding the data that were not observed. However, once a set of assumptions are made to obtain the estimate, the appropriate variance that corresponds with the estimate and assumptions should be used to test hypotheses.

In a randomized trial, the approach used to handle missing data should be specified in advance. Rather than simply assuming a model that would result in the most conservative estimated intervention differences, such as strategy 4, advice regarding potential missing data mechanisms should be sought from experienced interventionists. Additionally, individuals that may be involved in the potential public health policy implications of results from the trial should be consulted regarding possible interpretations in the presence of different missing data assumptions. Through advice from these individuals, and experience gained from other similar randomized trials, characterizations of missing data mechanisms that are more realistic than worst case scenarios can possibly be developed.

ACKNOWLEDGEMENTS

The authors would like to acknowledge support from NIH grants N01HC45138 and 5P60AG10 484, and thank the referees for the many useful comments.

REFERENCES

1. Gould AL. A new approach to the analysis of clinical drug trials with withdrawals. *Biometrics* 1980; **36**:721–727.
2. Pledger GW. Compliance in clinical trials: impact on design, analysis and interpretation. In *Compliance in Epilepsy* (*Epilepsy Res. Suppl. 1*), Schmidt D, Leppik IE (eds). Elsevier Science Publishers B.V.: The Netherlands, 1988; 125–133.
3. Wittes J, Lakatos E, Probstfeld J. Surrogate endpoints in clinical trials: cardiovascular diseases. *Statistics in Medicine* 1989; **8**:415–425.
4. Shih WJ, Quan H. Testing for treatment differences with dropouts present in clinical trials – A composite approach. *Statistics in Medicine* 1997; **16**:1225–1239.
5. Tilley BC, Alarcon GS, Heyse SP, Trentham DE, Neuner R, Kaplan DA, Clegg DO, Leisen JC, Buckley L, Cooper SM, Duncan H, Pillemer SR, Tuttleman MS, Fowler SE for the MIRA Trial Group. Minocycline in rheumatoid arthritis: a 48-week, double-blind, placebo-controlled trial. *Annals of Internal Medicine* 1995; **122**:81–89.
6. Tilley BC, Marler J, Geller NL, Lu M, Legler J, Brott T, Lyden P, Grotta J for the National Institute of Neurological Disorders and Stroke (NINDS) rt-PA Stroke Trial Study Group. Use of a global test for multiple outcomes in stroke trials with application to the National Institute of Neurological Disorders and Stroke *t*-PA Stroke trial. *Stroke* 1996; **27**:2136–2142.
7. The Writing Group for the DISC Collaborative Research Group. Efficacy and safety of lowering dietary intake of fat and cholesterol in children with elevated low-density lipoprotein cholesterol: The Dietary Intervention Study in Children (DISC). *Journal of the American Medical Association* 1995; **273**:1429–1435.
8. The Post Coronary Artery Bypass Graft Trial Investigators. The effect of aggressive lowering of low-density lipoprotein cholesterol levels and low-dose anticoagulation on obstructive changes in saphenous-vein coronary-artery bypass grafts. *New England Journal of Medicine* 1997; **336**(3):153–162.
9. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. Wiley: New York, 1987.
10. Lubson J. Exercise testing as outcome in congestive heart failure trials: design considerations when interpreting results. *Drugs* 1994; **47**(4):25–30.
11. Goetghebeur EJT, Shapiro SH. Analysing non-compliance in clinical trials: Ethical imperative or mission impossible? *Statistics in Medicine* 1996; **15**:2813–2826.
12. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley: New York, 1987.
13. Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis (with discussions). *Applied Statistics* 1994; **43**:49–93.
14. Grieve AP. Discussion of Diggle, P. and Kenward, MG. 'Informative drop-out in longitudinal data analysis'. *Applied Statistics* 1994; **43**:74–75.
15. Matthews JNS. Discussion of Diggle, P. and Kenward, MG. 'Informative drop-out in longitudinal data analysis'. *Applied Statistics* 1994; **43**:73–74.
16. Little RJA. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 1993; **88**(421):125–134.
17. Ettinger WH, Burns R, Messier SP, Applegate W, Rejeski WJ, Morgan T, Shumaker S, Berry MJ, O'Toole M, Monu J, Craven T. A randomized trial comparing aerobic exercise and resistance exercise with a health education program in older adults with knee osteoarthritis. *Journal of the American Medical Association* 1997; **277**:25–31.

