CLINICAL TRIALS TUTORIAL Clinical Trials 2005; 2: 379–386

Direct likelihood analysis versus simple forms of imputation for missing data in randomized clinical trials

Caroline Beunckens[a], Geert Molenberghs[a] and Michael G Kenward[b]

Background In many clinical trials, data are collected longitudinally over time. In such studies, missingness, in particular dropout, is an often encountered phenomenon.
Methods We discuss commonly used but often problematic methods such as complete case analysis and last observation carried forward and contrast them with broadly valid and easy to implement direct-likelihood methods. We comment on alternatives such as multiple imputation and the expectation-maximization algorithm.
Results We apply these methods in particular to data from a study with continuous outcomes. The outcomes are modelled using a general linear mixed-effects model. The bias with CC and LOCF is established in the case study and the advantages of the direct-likelihood approach shown.
Conclusions We have established formal but easy to understand arguments for a shift towards a direct-likelihood paradigm when analysing incomplete data from longitudinal clinical trials, necessitating neither imputation nor deletion. *Clinical Trials* 2005; **2**: 379–386. www.SCTjournal.com

Introduction

The randomized controlled trial (RCT) is often used to establish a causal effect of a new treatment on a response. Randomization permits valid causal inferences and effect estimation, allowing differences in clinical outcome to be ascribed to differences in treatment. This requires that investigators should follow two rules: first, they must define the rules that will govern random allocation; and, secondly, that allocation should be maintained, and all subsequent measurements of the patients should be complete and equal throughout the study.

However, in practice, this paradigm is jeopardized in two important ways. First, some patients may not receive the treatment as planned in the study protocol. One important reason subjects deviate from the planned schedule is due the occurrence of side effects. In some cases, patients may gain access to the treatment used in the other treatment arm(s). Hence, while patients remain on study, they do not follow the treatment regimen. To deal with this problem of nonadherence, the intention-to-treat (ITT) principle is often invoked [1–4]. ITT refers to an analysis that includes all randomized patients in the group to which they were randomly assigned, regardless of the appropriateness of their eligibility determination, of the treatment they actually received, or of deviation from the protocol. ITT analyses are primarily intended to address nonadherence problems that do not generate missing data, and it will not be dealt with further here [5].

A second threat to the validity produced by randomization is the departure of patients from a study or the failure to measure all relevant covariate and outcome information on a patient, producing missing (or "incomplete") data. Regardless of the cause of the missing information, inappropriate analysis of such data can lead to bias in the estimate of the treatment effect or its variance.

In the RCT setting, a commonly used method to analyse longitudinal data with nonresponse is based on setting a subject's missing response equal

[a]Center for Statistics, Limburgs Universitair Centrum, Diepenbeek, Belgium, [b]London School of Hygiene & Tropical Medicine, London, UK
Author for correspondence: Geert Molenberghs, Center for Statistics, Limburgs Universitair Centrum, Universitaire Campus, Building D, B-3590 Diepenbeek, Belgium. E-mail: geert.molenberghs@luc.ac.be

10.1191/1740774505cn119oa

to their last observed response (last observation carried forward, LOCF). Even though there have been many criticisms of this method [6, 7], it is still frequently seen in the context of regulated clinical studies. Other methods to deal with missing outcome data include complete case analysis (CC), available case analysis (AC) or simple forms of imputation. These are often employed without exploring the influence the assumptions underlying these methods might have on the final results. Several authors have written about this, a relatively early account being given provided by Heyting *et al.* [8]. Mallinckrodt *et al.* [9, 10] and Lavori [11] propose direct likelihood and multiple-imputation methods, respectively, to deal with incomplete longitudinal data. Siddiqui and Ali [12] compare direct likelihood and LOCF methods.

Following several authors [13–16], we will show how analysing the data as if it were complete after carrying the last observation forward, or analysing only those subjects with complete data, are an unscientific approaches. This message will be illustrated in a number of simple settings and a real example. We will demonstrate that there are a number of likelihood-based methods which can be implemented with standard software that make assumptions underlying the missing-data handling transparent, and allow the sensitivity of the conclusions to these assumptions to be assessed and reported.

Missingness mechanisms

We will focus on dropout from a RCT. The process or mechanism that caused the outcomes to be missing, is called the missingness, or nonresponse process. We discuss the three broad causes of missing data introduced by Rubin [17].

The missingness process is said to be missing completely at random (MCAR) if the data are missing for reasons unrelated to the response or to characteristics of individuals. For example, a subject may move, their data may be lost due to an administrative mix up, or they may simply tire of participating in the study. However, the reasons for missingness are not always easy to ascertain. For example, if a subject withdrew because they experienced a car accident, their outcome data might be considered MCAR, but perhaps should not be if the subject's treatment could have affected their ability to drive. As will be discussed in subsequent sections, methods like LOCF and CC are based on extremely strong assumptions about missingness and even the strong MCAR assumption does not suffice to guarantee that an LOCF analysis is valid.

Data are missing at random (MAR) if the cause of dropout is allowed to depend on the subject's observed data, but not on their post-dropout, or unobserved, responses. Such scenarios are more common than MCAR. An example of MAR is a trial in which subjects are removed if their response has exceeded a prespecified limit. Alternatively, subjects may quit the trial if they are either doing much better or significantly worse.

The MAR assumption implies that future behaviour for those who share the same past measurements and covariate values is on average identical whether or not they drop out. This enables treatment effect to be estimated in longitudinal models without simultaneously modelling the cause of dropout.

Under MAR, valid inference can be obtained through a likelihood-based analysis, without the need for modelling the dropout process. As a consequence, one can simply use, for example, linear or generalized linear mixed models [13], without additional complication or effort. We will show that such an analysis not only enjoys much wider validity than the simple methods but in addition is easy to conduct, without additional data manipulation using such tools as the SAS procedures MIXED or NLMIXED, HLM4.0, the SPSS procedure MIXED, the SPlus functions lme and nlme and MLwiN, to name a few. In the standard linear mixed model case, all of these can handle incomplete data, and provide the same conclusions since the same estimates result. For non-Gaussian longitudinal data, the situation is somewhat more complex [14]. There is no reason to use ad hoc methods when direct likelihood analyses can be implemented with standard software.

If the cause of missing data is neither MCAR nor MAR, the data is missing not at random (MNAR). In the most general setting, the cause of a subject's dropout depends on their post-dropout, unobserved, responses, even after allowing for the information of the observed data. In this case, the reason for dropout should be modelled simultaneously with the response. An example of MNAR data would be a subject who had been doing well until midway in a trial but relapsed after the last observed visit and was lost to follow-up.

In the case of likelihood-based estimation, given that the parameters defining the measurement process are independent of the parameters defining the missingness process, missingness is said to be ignorable if it arises from an MCAR or MAR process. Missingness is nonignorable if it arises from an MNAR process.

In realistic settings, the reasons for dropout are varied and it is therefore difficult to fully justify on *a priori* grounds the assumption of MAR. Further, since it is not possible to test for MNAR against MAR, one should always be open to the possibility that the data are MNAR. To explore the impact of

deviations from the MAR assumption on the conclusions, one should ideally conduct a sensitivity analysis, within which models for the MNAR process can play a major role [13].

Methods in common use

Complete case analysis

A complete case analysis includes only those cases for analysis, for which all measurements – covariates and outcomes – were recorded [13, 14, 18]. This method has obvious advantages. It is very simple to describe and since the data structure looks like it resulted from a complete experiment, standard statistical software can be used without additional work. Further, since the entire estimation is done on the same subset of completers, there is a common basis for inference. Unfortunately, the method suffers from severe drawbacks. First, there can be a substantial loss of information, with adverse effects on precision and power, even if the frequency of missing data for single variables low. Further, such an analysis will only be representative for patients who remain on study and have complete data. A complete case analysis can have a role as an auxiliary analysis, especially if it relates to a scientific question. A final important issue about a complete case analysis is that it is only valid when the missingness mechanism is MCAR. Severe bias can result when the missingness mechanism is MAR but not MCAR. This bias can be positive or negative, as illustrated by Molenberghs *et al.* [17].

Last observation carried forward

A method that has received a lot of attention [9, 10, 12–14, 18] is last observation carried forward (LOCF). In the LOCF method, whenever a value is missing, the last observed value is substituted. For the LOCF approach, the MCAR assumption is necessary but not sufficient. This approach further assumes that subjects' responses would have been unchanged from the last observed value to the endpoint of the trial. These conditions seldom hold [13, 14]. In a clinical trial setting, one might believe that the response profile changes as soon as a patient goes off treatment and plateaus thereafter. However, the LOCF constant profile assumption is even stronger. Therefore, carrying observations forward may bias estimates of treatment effects in either direction and will underestimate the associated standard errors [8, 11, 12, 13, 16, 19–21]. This method artificially increases the amount of information in the data by treating imputed and actually observed values on equal footing.

Despite its shortcomings, LOCF has been the longstanding method of choice for the primary analysis in clinical trials because of its simplicity, ease of implementation with standard software, and the belief that the potential bias from carrying observations forward leads to a "conservative" analysis. An analysis is deemed conservative when the treatment effect estimated is smaller in absolute value than the true one. However, examples of anti-conservative effect of LOCF are common [22–24], meaning an LOCF analysis can create the appearance of a treatment effect when none exists.

Available case analysis

In a traditional available case analysis (AC) [18], estimators are based on the subjects who have complete information available for a specific analysis, a subset that can change when different covariates or time points are considered. For example, a collection of such analyses could be the treatment-specific means at a series of designated measurement times. With increasing dropout over time, means later in the study would be calculated using fewer subjects than earlier means.

Direct likelihood approaches when data are incomplete

An alternative approach for handling missing data in an RCT is to use methods that are valid under the weaker MAR assumption [13, 14, 18], instead of the methods discussed in previous section for which the MCAR assumption, and more, is needed. Note that methods valid under MAR are also valid if data are MCAR, while the reverse does not hold.

Laird and Ware [13,25] proposed, for continuous outcomes, likelihood-based mixed-effects models, which are valid under the MAR assumption. For longitudinal studies with missing data, a mixed model only requires that missing data are MAR. These mixed-effects models permit the inclusion of subjects with missing values at some time points (both dropout and intermittent missingness).

For clarity, let us introduce the general linear mixed-effects model:

$$\begin{cases} Y_i = X_i\beta + Z_i b_i + \varepsilon_i \\ b_i \sim N(0, D) \\ \varepsilon_i \sim N(0, \Sigma_i) \\ b_1, \ldots, b_N, \varepsilon_1, \ldots, \varepsilon_N \text{ independent} \end{cases}$$

where Y_i is the n_i dimensional response vector for patient i, containing the outcomes at n_i various measurement occasions, $1 \leq i \leq N$, N is the number of patients, X_i and Z_i are $(n_i \times p)$ and $(n_i \times p)$

dimensional matrices of known covariates, β is the p dimensional vector containing the fixed effects, b_i is the q dimensional vector containing the random effects, and ε_i is a n_i dimensional vector of residual components, combining measurement error and serial correlation. Finally, D is a general $(q \times q)$ covariance matrix with (i, j) element $d_{ij} = d_{ji}$ and Σ_i is a $(n_i \times n_i)$ covariance matrix which generally depends on i only through its dimension n_i, i.e., the set of unknown parameters in Σ_i will not depend upon i.

This likelihood-based MAR analysis is also termed likelihood-based ignorable analysis, or, as we will term it in the remainder of this paper, a direct likelihood analysis. In the literature, names for these methods vary, and include hierarchical models, random-effects models, and random-coefficient models.

In a direct likelihood analysis, the observed data are used without deletion nor imputation. In doing so, appropriate adjustments, i.e., valid under MAR, are made to parameters at times when data are incomplete, due to the within-patient correlation. Even when interest lies in a comparison between the two treatment groups at the last measurement time, such a full longitudinal analysis is a good approach, since the fitted model can be used as the basis for inference.

Let us take a look at an artificial but revealing example contrasting this approach with the ones discussed earlier. Figure 1, displays the results of the traditional MCAR methods – complete case, available case and LOCF – with the result of an MAR method. In this example, the mean response is supposed to be a linear function of the variable on the abscissa. For patients with incomplete data and others with complete observations, the slope is the same, but intercepts differ. We assume that patients with incomplete observations drop out half way through the study (time point 5) upon reaching a certain level of the response – an MAR missingness mechanism. Using a method valid under the MAR assumption, the analysis would yield the correct mean profile, a straight line centred between the mean profiles of the completers and noncompleters. If one performed a complete case analysis, the fitted profile will coincide with the mean profile of the complete cases (bold line). Under LOCF, Figure 1 (dashed line) shows the data that are observed (bold dashed line), a progressively increasing underestimate of the true mean. Finally, this figure shows how the AC approach (bold dash-dot line) can produce anomalous results in this situation; the trajectory becomes discontinuous at time point 5, with a mean identical to those who continue beyond that point. All of the MCAR methods produce incorrect results under this simple but plausible scenario.

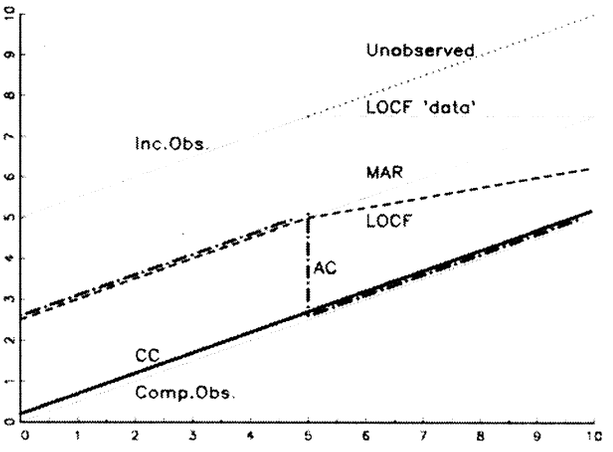


Figure 1 Artificial example of a study where subjects drop out at time point 5 (x axis) after reaching a certain level of response (y axis). The dataset is composed of those with complete observations (bottom thin line) and those who have incomplete observations (top thin line). The estimated trajectories of this cohort using different analytic approaches to handle the incomplete data are shown. (MAR – middle bold line, LOCF – middle bold dashed line, AC – bold dash-dotline, CC bottom bold line. The MAR line represents the correct result. The upper two lines show what would have happened to the subjects who dropped out; their actual unobserved average (dotted line) and the average assumed by the LOCF approach (long dash line)

In many RCT settings the repeated measures are balanced, in the sense that a common (and often limited) set of measurement times is considered for all subjects, which allows the *a priori* specification of a "saturated" model, such as, for example, a full group-by-time interaction for the fixed effects combined with an unstructured covariance matrix. Such a model specification is termed mixed-effects model repeated-measures analysis (MMRM) by Mallinckrodt *et al.* [20, 21]. MMRM is a particular form of a linear mixed model, relevant for confirmatory clinical trials, fitting within the direct likelihood paradigm. This direct likelihood MMRM analysis is equivalent to a classical MANOVA analysis when data are complete, but valid when they are incomplete.

Therefore, direct likelihood methods are a very promising alternative to the simple methods described in the previous section. When a relatively large number of measurements is made within a single subject, the full power of random effects modelling can be used [13].

It is sometimes stated that LOCF is a preferred approach when the ITT principle [2, 6] is adhered to since data on all patients randomized can be used. Direct likelihood methods also use information on all subjects; including information from early dropouts, while avoiding the much stronger assumptions required to make LOCF valid. Thus, the direct likelihood method is a sensible approach under ITT.

There are a number of alternatives to direct likelihood. One of these is multiple imputation (MI) [18, 26]. The MI method involves constructing a number of complete datasets from an incomplete one by drawing from the conditional distribution of the unobserved outcomes, given the observed ones. These datasets are then analysed and the results combined to produce inferences. Verbeke and Molenberghs [13] discuss the method in the context of continuous longitudinal data. Molenberghs and Verbeke [14] illustrate how the SAS procedures MI and MIANALYZE can be used in this context. Multiple imputation is valid under the same conditions as direct likelihood, and therefore does not suffer from the problems encountered in most single imputation methods. However, there are a number of situations where multiple imputation are particularly useful. For example, when outcomes as well as covariates are missing then multiple imputation is a sensible route. The method is also useful when several analyses, perhaps conducted by different analysts, have to be done on the same set of incomplete data. In such a case, all analyses could start from the same set of multiply-imputed sets of data and enhance comparability. Another method that has seen a number of applications is the expectation-maximization (EM) algorithm [18, 27]. Broadly speaking, the algorithm is a general method to fit a likelihood to incomplete data. When used as an alternative to the direct likelihood method used in this paper, it should give the exact same estimates, but computations are more difficult. Verbeke and Molenberghs [13, 14] show how the method can be used in SAS. For most standard longitudinal clinical trial settings, we recommend that direct likelihood be the first choice.

Estimates in case of two measurements

Using the simple setting of two repeated follow-up measures, the first of which is always observed while the second can be missing, we establish some properties of the LOCF and CC estimation procedures, both assuming MCAR, as well as the estimation procedure when the missingness mechanism is assumed to be MAR.

Let us assume each subject $i = 1, \ldots, N$ in the study is to be measured on two occasions. The responses are grouped in a two-component vector (Y_{i1}, Y_{i2}). Assume a linear mixed model, with constant mean for both time points, and an unstructured variance-covariance matrix:

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ & \sigma_2^2 \end{pmatrix}\right]$$

Table 1 Estimates of $\hat{\mu}_1$ and $\hat{\mu}_2$: MAR (all incomplete data under MAR), CC (complete cases only), versus LOCF (LOCF imputed data)

	MAR	CC	LOCF
$\hat{\mu}_1$	$\bar{y}_1^{(n)}$	$\bar{y}_1^{(r)}$	$\bar{y}_1^{(n)}$
$\hat{\mu}_2$	$\bar{y}_2^{(r)} + \frac{n-r}{n}\hat{\beta}_1\left(\bar{y}_1^{(n-r)} - \bar{y}_1^{(r)}\right)$	$\bar{y}_2^{(r)}$	$\bar{y}_1^{(r)} + \frac{r}{n}\left(\bar{y}_1^{(r)} - \bar{y}_1^{(r)}\right)$

Further, assume the first r of these subjects complete the study, while for the remaining ones only the first measurement is observed. Conditional on the first observation, the second measurement will also be normally distributed with mean linearly related to the value of the first observation y_{i1} and variance $\sigma^2_{2|1}$:

$$Y_{i2}|Y_{i1} = y_{i1} \sim N(\beta_0 + \beta_1 y_{i1}, \sigma^2_{2|1})$$

where

$$\begin{cases} \beta_1 = \rho\dfrac{\sigma_2}{\sigma_1} \\ \beta_0 = \mu_2 - \beta_1\mu_1 = \mu_2 - \rho\dfrac{\sigma_2}{\sigma_1}\mu_1 \\ \sigma^2_{2|1} = \sigma_2^2(1-\rho^2) \end{cases}$$

The estimates of the mean parameters, $\hat{\mu}_1$ and $\hat{\mu}_2$, using either the CC or LOCF method, or a method valid under the MAR assumption, are listed in Table 1.

Note that, under LOCF, a correction is taking place without any adjustment for the correlation between the two measurements, whereas it is only correct in the unlikely case of correlation exactly equal to one and means constant across measurement occasions. Thus, LOCF would be inappropriate, and dramatically so in the zero correlation situation. Under CC, the means at both measurement times are incorrect, even thought there is no need for a correct at the first one. The direct likelihood method uses the difference between the means for complete and incomplete observations at time one, modifed by the correlation between the two measurement occasions, to correct the mean at the second occasion.

Example: growth data

As an example, we use the orthodontic growth data, introduced by Pothoff and Roy [28], and having characteristics of a RCT. These data contain growth measurements for 11 girls and 16 boys. For each subject, the distance in millimetres from the centre

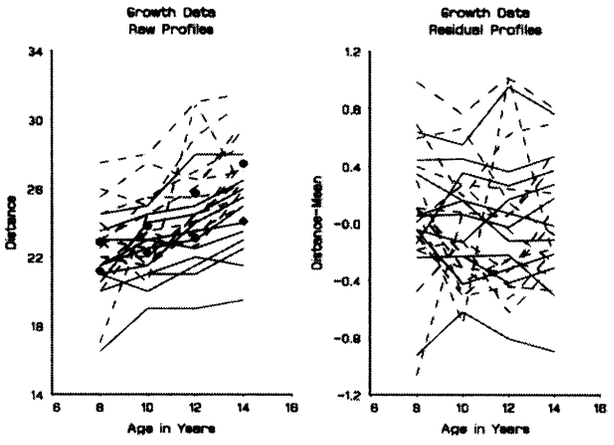


Figure 2 Raw and residual profiles for the complete growth data set (Girls are indicated with solid lines. Boys are indicated with dashed lines.)

of the pituitary to the maxillary fissure was recorded at ages 8, 10, 12 and 14. Figure 2 presents the 27 individual profiles. Little and Rubin [29] deleted nine of the $[(11 + 16)\times 1]$ measurements, rendering nine incomplete subjects. Deletion is confined to the age 10 measurements. Little and Rubin [18] describe the mechanism as subjects with a low value at age eight being more likely to have a missing value at age 10. The advantage of this example is that we have the complete data set, i.e., the original data, as well as the incomplete data available.

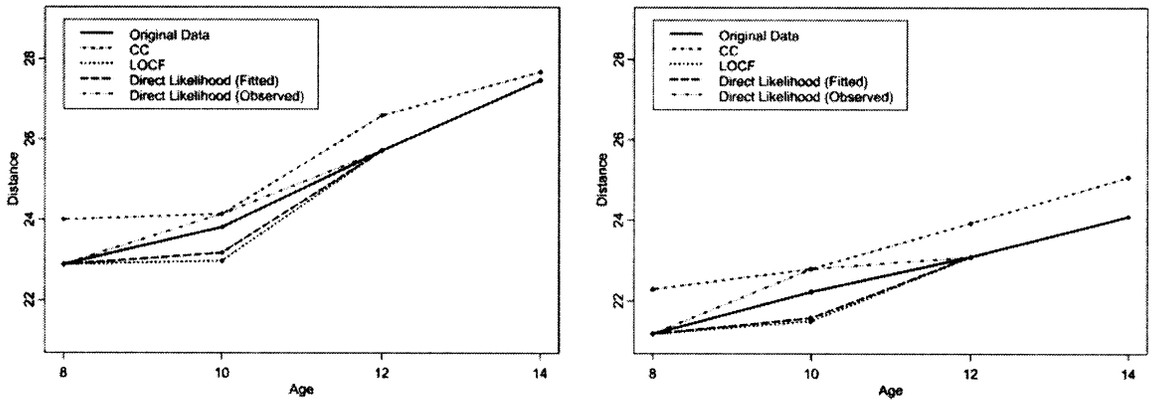
The simple methods and direct likelihood method from previous sections are now compared using the growth data. We analysed the original data, next to the CC data, the LOCF data, and the incomplete data as such. For this purpose, a linear mixed model is used, assuming unstructured mean, i.e., assuming a separate mean for each of the eight age×sex combinations, together with an unstructured covariance structure, and using maximum likelihood (ML) as well as restricted maximum likelihood (REML). The mean profiles of the linear mixed model using maximum likelihood for all four data sets are given in Figure 3 for boys and girls separately.

Next to this longitudinal approach, we will consider a MANOVA analysis and an ANOVA analysis per time point. For all these analyses, Table 2 shows the estimates and standard errors for boys at age eight and 10, for the original data and all available incomplete data, as well as for the CC and the LOCF data.

First, we consider the group means for both sex groups for the original data set in Figure 3, i.e., we observe relatively straight lines both in left and right panel. Clearly, there seems to be a linear trend in both profiles.

In a complete case analysis of the growth data, the nine subjects which lack one measurement are deleted, resulting in a working data set with 18 subjects. This implies that 27 available measurements will not be used for analysis, a severe penalty on a relatively small data set. Observing the profiles for the CC data set in Figure 3, all group means increase relative to the original data set but mostly so at age eight. The net effect is that the profiles overestimate the average length.

For the LOCF data set, the nine subjects that lack a measurement at age 10 are completed by imputing the age eight value. It is clear that this procedure will affect the linear but nonconstant trend model found for the original data set. Indeed, the



(a) Boys (b) Girls

Figure 3 Profiles for the growth data using the original data, CC, LOCF and direct likelihood method

Table 2 Comparison of analyses

Method	Boys at age 8	Boys at age 10
Original data		
Direct likelihood, ML	22·88 (0·56)	23·81 (0·49)
Direct likelihood, REML	22·88 (0·58)	23·81 (0·51)
MANOVA	22·88 (0·58)	23·81 (0·51)
ANOVA per time point	22·88 (0·61)	23·81 (0·53)
All available incomplete data		
Direct likelihood, ML	22·88 (0·56)	23·17 (0·68)
Direct likelihood, REML	22·88 (0·58)	23·17 (0·71)
MANOVA	24·00 (0·48)	24·14 (0·66)
ANOVA per time point	22·88 (0·61)	24·14 (0·74)
Complete case analysis		
Direct likelihood, ML	24·00 (0·45)	24·14 (0·62)
Direct likelihood, REML	24·00 (0·48)	24·14 (0·66)
MANOVA	24·00 (0·48)	24·14 (0·66)
ANOVA per time point	24·00 (0·51)	24·14 (0·74)
Last observation carried forward analysis		
Direct likelihood, ML	22·88 (0·56)	22·97 (0·65)
Direct likelihood, REML	22·88 (0·58)	22·97 (0·68)
MANOVA	22·88 (0·58)	22·97 (0·68)
ANOVA per time point	22·88 (0·61)	22·97 (0·72)

imputation procedure forces the means at ages eight and 10 to be very similar, thereby destroying the linear relationship. Hence, a simple, intuitively appealing interpretation of the trends is made impossible.

In case of direct likelihood, we now see two profiles. One for the observed means (based on the available sample at each point in time) and one for the fitted means. These two coincide at all ages except age 10. At first sight, this is confusing because our model is a seemingly saturated one. However, the well-known fact that a saturated time-by-treatment group model reproduces the observed means is true only when the data are balanced, in the sense that all subjects have measurements at exactly the same times. Missingness disturbs this designed balance. This is a strength of the likelihood method, since it takes a correction into account, based on the observed data of a subject with incomplete data (see Table 1). As mentioned earlier, the complete observations at age 10 are those with a higher measurement at age eight. Due to the within-subject correlation, they are the ones with a higher measurement at age 10 as well, and therefore the fitted likelihood model corrects in the appropriate direction.

As an aside, note that in the case of direct likelihood, the observed average at age 10 coincides with the CC average, while the fitted average does not coincide with anything else. Indeed, if the model specification is correct, then a direct likelihood analysis produces a consistent estimator for the average profile, as if nobody had dropped out. This effect might be obscured in small data sets due to large variability. In spite of the small-sample behaviour encountered here, the validity under MAR and the ease of implementation are good arguments that favor this direct likelihood analysis over other techniques.

Let us now compare the different methods by looking at Table 2, which shows the estimates and standard errors for boys at age eight and 10, for the original data and all available incomplete data, as well as for the CC data and the LOCF data.

Table 2 shows some interesting features. In all four cases, a CC analysis gives an upward biased estimate, for both age groups. This is obvious, since the complete observations at age 10 are those with a higher measurement at age eight, as we have seen before. The LOCF analysis gives a correct estimate for the average outcome for boys at age eight. This is not surprising since there were no missing observations at this age.

As noted before, the estimate for boys of age 10 is biased downwards. When the incomplete data are analysed, we see from Table 2 that direct likelihood produces good estimates. The MANOVA and ANOVA per time point analyses give an overestimation of the average of age 10, as in the CC analysis. Further, the MANOVA analysis also yields an overestimation of the average at age eight, again the same as in the CC analysis.

In complete sets of data, direct likelihood, especially with the REML estimation method, is identical to MANOVA (see Table 2). Given the classical robustness of MANOVA, and its close agreement with ANOVA per time point, this provides an additional basis for using direct likelihood, which is not as assumption-driven as is sometimes believed. This, in addition to the validity of direct likelihood under MAR (and hence its divergence from MANOVA and ANOVA for incomplete data) provides a strong justification for the direct likelihood method.

Conclusion

A direct likelihood analysis uses all available information, without the need either to delete nor to impute measurements or entire subjects. It is theoretically justified whenever the missing data mechanism is MAR, a less restrictive and more realistic assumption than MCAR, which is necessary (but not always sufficient) for simple analyses (AC, CC, LOCF). There is no distortion in the statistical information, since observations are neither removed (such as in CC analysis) nor added (such as in LOCF analysis). To perform a direct likelihood analysis, standard software can be applied, and no additional programming is involved.

Acknowledgements

Caroline Beunckens and Geert Molenberghs gratefully acknowledge support from *Fonds Wetenschappelijk Onderzoek-Vlaanderen* Research Project G.0002.98 "Sensitivity Analysis for Incomplete and Coarse Data" and from Belgian IUAP/PAI network "Statistical Techniques and Modeling for Complex Substantive Questions with Complex Data".

References

1. **Hill AB.** *Principles of medical statistics*, seventh edition. London: The Lancet, 1961.
2. **Schwartz D, Lellouch J.** Explanatory and pragmatic attitudes in therapeutic trials. *Journal of Chronic Diseases* 1967; **20**: 637–48.
3. **Armitage P.** Attitudes in clinical trials. *Statistics in Medicine* 1998; **17**: 2675–83.
4. **McMahon AD.** Study control, violators, inclusion criteria and defining explanatory and pragmatic trials. *Statistics in Medicine* 2002; **21**: 1365–76.
5. **Fischer K, Goetghebeur E.** Structural mean effects of noncompliance: estimating interaction with baseline prognosis and selection effects. *Journal of the American Statistical Association* 2004; **99**: 918–28.
6. **Pocock SJ.** *Clinical trials: a practical approach.* Chichester: John Wiley, 1983.
7. **Pocock SJ.** Advances in Biometry. *Clinical trials: a statistician's perspective.* In Armitage P, David HA eds. New York: Wiley, 1996.
8. **Heyting A, Tolboom J, Essers J.** Statistical handling of dropouts in longitudinal clinical trials. *Statistics in Medicine* 1992; **11**: 2043–61.
9. **Mallinckrodt CH, Clairk WS, Carroll RJ, Molenberghs G.** Assessing response profiles from incomplete longitudinal clinical trial data under regulatory considerations. *Journal of Biopharmaceutical Statistics* 2003a; **13**: 179–90.
10. **Mallinckrodt CH, Sanger TM, Dube S** *et al.* Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. *Biological Psychiatry* 2003b; **53**: 754–60.
11. **Lavori PW, Dawson R, Shera D.** A multiple imputation strategy for clinical trials with truncation of patient data. *Statistics in Medicine* 1995; **14**: 1913–25.
12. **Siddiqui O, Ali MW.** A comparison of the random-effects pattern mixture model with last observation carried forward (LOCF) analysis in longitudinal clinical trials with dropouts. *Journal of Biopharmaceutical Statistics* 1998; **8**: 545–63.
13. **Verbeke G, Molenberghs G.** *Linear mixed models for longitudinal data.* New York: Springer, 2000.
14. **Molenberghs G, Verbeke G.** *Models for discrete longitudinal data.* New York: Springer, 2005.
15. **Ware JH.** Interpreting incomplete data in studies of diet and weight loss. *New England Journal of Medicine* 2003; **348**: 2136–37.
16. **Molenberghs G, Thijs H, Jansen I** *et al.* Analyzing incomplete longitudinal clinical trial data. *Biostatistics* 2004; **5**: 445–64.
17. **Rubin DB.** Inference and missing data. *Biometrika* 1976; **63**: 581–92.
18. **Little RJA, Rubin DB.** *Statistical analysis with missing data.* New York: John Wiley & Sons, 2002.
19. **Gibbons RD, Hedeker D, Elkin I** *et al.* Some conceptual and statistical issues in analysis of longitudinal psychiatric data. *Archives of General Psychiatry* 1993; **50**: 739–50.
20. **Mallinckrodt CH, Clark WS, Stacy RD.** Type I error rates from mixed-effects model repeated measures versus fixed effects analysis of variance with missing values imputed via last observation carried forward. *Drug Information Journal* 2001a; **4**: 1215–25.
21. **Mallinckrodt CH, Clark WS, Stacy RD.** Accounting for dropout bias using mixed-effects models. *Journal of Biopharmaceutical Statistics* 2001b; **1 & 2**: 9–21.
22. **Mallinckrodt CH, Watkin JG, Molenberghs G, Carroll RJ.** Choice of the primary analysis in longitudinal clinical trials. *Pharmaceutical Statistics* 2004; **3**: 161–69.
23. **Little RJA, Yau L.** Intent-to-treat analysis in longitudinal studies with drop-outs. *Biometrics* 1996; **52**: 1324–33.
24. **Liu G, Gould AL.** Comparison of alternative strategies for analysis of longitudinal trials with dropouts. *Journal of Biopharmaceutical Statistics* 2002; **12**: 207–26.
25. **Laird NM, Ware JH.** Random effects models for longitudinal data. *Biometrics* 1982; **38**: 963–74.
26. **Rubin DB.** *Multiple imputation for nonresponse in surveys.* New York: John Wiley & Sons, 1987.
27. **Dempster AP, Laird NM, Rubin DB.** Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B* 1977; **39**: 1–38.
28. **Potthoff RF, Roy SN.** A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* 1964; **51**: 313–26.
29. **Little RJA, Rubin, DB.** *Statistical analysis with missing data.* New York: John Wiley, 1987.