Modern statistical methods for handling missing repeated measurements in obesity trial data: beyond LOCF

G. L. Gadbury[1], C. S. Coffey[2] and D. B. Allison[3]

[1]Department of Mathematics and Statistics, University of Missouri – Rolla, Rolla, MO, USA; [2]Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL, USA; [3]Section on Statistical Genetics, Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL, USA

Received 31 March 2003; revised 22 May 2003; accepted 26 May 2003

Address reprint requests to: GL Gadbury, Department of Mathematics and Statistics, University of Missouri – Rolla, Rolla, MO 65409, USA. E-mail: gadburyg@umr.edu

Summary

This paper brings together some modern statistical methods to address the problem of missing data in obesity trials with repeated measurements. Such missing data occur when subjects miss one or more follow-up visits, or drop out early from an obesity trial. A common approach to dealing with missing data because of dropout is 'last observation carried forward' (LOCF). This method, although intuitively appealing, requires restrictive assumptions to produce valid statistical conclusions. We review the need for obesity trials, the assumptions that must be made regarding missing data in such trials, and some modern statistical methods for analysing data containing missing repeated measurements. These modern methods have fewer limitations and less restrictive assumptions than required for LOCF. Moreover, their recent introduction into current releases of statistical software and textbooks makes them more readily available to the applied data analyses.

Keywords: Clinical trial, ignorable, imputation, missing data, mixed model, random effects

obesity reviews (2003) **4**, 175–184

Introduction

A number of studies have reported on relationships between obesity and health risks (1,2), a reduced quality of life (3–6), an added economic burden on society (7–10), and increased mortality (11–13). The last relationship between obesity and mortality has prompted recent discussions regarding the implications of intentional vs. unintentional weight loss on mortality (14,15). However, the prevalence of obesity has been increasing worldwide (2,16), and the prevalence of childhood obesity is also increasing (17) prompting added concerns over the predictive nature of this condition for ill-health later in life (18), and the initiation of studies to evaluate factors linked to obesity in children (19–21).

Given the prevalence and deleterious consequences of obesity described above, and the fact that currently available methods for treating and preventing obesity are, at best, modestly successful (1,22–24), many investigators are studying new approaches to treat and prevent obesity (cf. 25–27). The randomized controlled clinical trial is thought to be the 'gold standard' for establishing a causal effect of a new treatment on a response (28). Conducting such trials to evaluate treatments for obesity offers special challenges, most notably, a requirement that the length of the trial be sufficient to confirm treatment efficacy not only for short-term use (29).

In obesity trials, statistical evaluation of a treatment's efficacy is usually complicated by missing observations because of dropouts (i.e. subjects who drop out of the clinical trial after some interim follow-up visit and do not return) or by missing observations because of subjects who miss one or more visits (even though they might complete the trial). For example, one study evaluating two doses of sibutramine vs. a placebo had 485 subjects randomized into three treatment arms, and 53% of subjects completed

the 12-month trial. Estimating a treatment effect at the final visit required a technique to handle the 226 missing observations at that visit, not to mention any additional missing data because of subjects who also missed interim monthly visits.

One technique that has been used to handle missing data in obesity trials with repeated measurements is 'last observation carried forward' (LOCF) (cf. 30,31). This technique replaces subjects' missing outcomes with outcomes observed on the previous visit. For example, if there were five follow-up visits in a clinical trial (denoted times 1, 2, …, 5) and a subject dropped out after the third visit, observed outcomes would be y_1, y_2, y_3 and outcomes y_4, y_5 would be missing. LOCF would set $y_5 = y_4 = y_3$. This method is easy to implement in practice and sometimes, but not necessarily, provides a conservative estimate of a treatment's effect (i.e. the effect of treatment is underestimated) (32). However, the assumption that a subject's outcome would remain constant can not only bias estimates of a treatment's effect, but also underestimate the true variability in the data and increase the probability of making a type 1 error above the desired target significance level (32). Another option, as easily implemented in practice as LOCF, is to analyse data from 'completers' only. In this rather extreme technique, the above subject who dropped out after the third visit would be omitted entirely from the analysis, thus ignoring potentially valuable information regarding the treatment (i.e. the information available in the observed y_1, y_2, y_3).

There is a large body of literature describing statistical methods to handle missing data and of which the associated theoretical properties have been well studied (theoretical properties of LOCF have been given limited attention). Although a number of these methods evolved from the need to analyse data from surveys, many have been extended and adapted to longitudinal data and one method in particular is specifically suited to such. The methods are relatively recent (from a historical perspective), the background literature on some still theoretical, and the software to implement them relatively recent and less trivial to implement than the more common methods available for complete and balanced designs (i.e. no missing data). The methods have been used for data analyses and reported in most applied statistical journals (33,34), but their adaptation into obesity trial data has been slow with a few exceptions (e.g. 27,35). The field of obesity research is not alone in dealing with the problems associated with missing data in longitudinal studies. Recent publications in the field of psychiatric research (36,37) illustrate that similar problems also plague other fields of research.

This paper evaluates some recent developments in statistics within the context of randomized trials to evaluate obesity treatments where repeated measurements are taken on subjects and dropout is an issue. We will not discuss mathematical details [this has been done elsewhere (e.g. 34,38,39)]. This material will inform researchers in obesity about other statistical techniques beyond LOCF for handling missing data from clinical studies. Ideally, there would be no missing data resulting from a clinical trial, and method to retain subjects in clinical trials is its own area of research (cf. 40), and beyond the scope of this paper. We also do not consider the topic of treatment compliance, that is, a subject who remains in a trial but does not comply with an assigned treatment. Further details on design of obesity trials and a discussion of some statistical issues have appeared elsewhere (41,42).

The missing data mechanism

The central question of many studies involving missing data has seemed to be 'what should be done about the missing outcomes, that is, should they be filled in via a technique like LOCF or should the entire cases be omitted from the analysis?' More problematic, however, is the often ignored process or mechanism that caused the outcomes to be missing in the first place. The first time that the theoretical aspects of this problem were rigorously considered in the statistical literature may have been Rubin (43), an article that began further initiatives that continue to this day (e.g. 44–46).

As an illustration, consider the example subject from above who enters an obesity trial with an initial 'baseline' visit and five scheduled follow-up visits. The subject's conceptual outcomes (e.g. change from baseline) may be denoted Y_1, Y_2, Y_3, Y_4, Y_5. Suppose, again, that the subject will complete the third follow-up visit and then drop out of the trial. Then Y_4, Y_5 are unobservable and resulting data will contain missing values for these two variables.

There is another variable for this subject, call it R, that defines the process that causes missing outcomes. This variable can be defined in different ways. For an arbitrary pattern of missing data, R could be defined as a vector of five ones and zeros. For example, $R = [1,1,0,0,1]$ would mean that a subject attended visits 1, 2, and 5 but missed visits 3 and 4. An alternative definition that we will adopt, for convenience, can be used for a monotone pattern of missingness, that is, once a subject drops out, he/she does not return for any remaining follow-up visit. In this case, R can be defined as the follow-up visit number before the subject dropped out. For the subject described earlier who dropped out after the third visit, $R = 3$, but the variable itself may take on integer values from 0 to 5, 0 meaning the subject never returned after the baseline visit and 5 meaning the subject completed the trial. Table 1 shows a monotone missingness pattern for a hypothetical set of four subjects. The modern statistical methods to be described, however, are still applicable with arbitrary patterns of missingness.

For each subject, R can take on values from 0 to 5 with, it is likely, different probabilities. The relationship between these probabilities and the outcome variable Y determines the extent to which this variable can be entirely ignored in a statistical analysis of a treatment's effect on values of Y. Let Y_{obs} be the set of observed changes from baseline in the outcome variable, and let Y_{mis} be the set of missing observations. Thus, for subject 1 in Table 1, $Y_{obs} = (Y_1, Y_2, Y_3)$ and $Y_{mis} = (Y_4, Y_5)$.

Suppose X denotes a set of other variables observed for a subject. Examples are a subject's baseline measurement of the outcome variable, Y_0, a subjects age at baseline, gender, race, etc., and may also include an indication of the treatment group to which a subject was assigned. If R is independent of (the probabilities for values of R do not depend on) $(Y_{obs}, Y_{mis}) = (Y_1, Y_2, Y_3, Y_4, Y_5)$ in Table 1, or any other variables X, then the missing values are said to be 'missing completely at random' (MCAR). A special case of MCAR occurs when R is independent of (Y_{obs}, Y_{mis}), but can depend on X. This special case has been termed 'covariate – dependent dropout' (44), that we denote MCAR2. If R is independent of the values that will be missing, that is, Y_{mis}, given other variables that are observed, that is, (Y_{obs}, X), then the missing values are said to be 'missing at random' (MAR). Missing data that are MAR or MCAR (or MCAR2) are said to result from a missing data mechanism that is ignorable. For data of this type, we do not have to specify a distribution for the pattern of missingness in order to obtain valid inferences regarding the primary parameters of interest. Furthermore, MCAR or MCAR2 may be testable using observed data by comparing distributions of observed variables across the patterns of missing data. That is, MCAR, MCAR2, and MAR are nested in that the former two may be empirically testable while MAR is not directly testable.

If the probabilities for values of R depend on Y_{mis}, and if this dependence cannot be eliminated by adjusting for other observed variables, then the missing data mechanism is not ignorable and missing data resulting from this are described as 'not missing at random' (NMAR). Table 2 summarizes the extent to which missing data mechanisms can be ignored in a statistical analysis. This is shown via the definition of a missing data mechanism denoted as a probability equation. So, for example, data that are MCAR result from $P(R = r1|2Y_{obs}, Y_{mis}, X) = P(R = r)$ which reads, 'the probability that the subject drops out at the rth follow-up visit, given other variables to be observed or to be missing and covariates, X, is equal to the probability that the subject drops out at the rth follow-up visit.' This suggests that the subject dropped out for reasons unrelated to the treatment, or how well they were or were not benefiting from the treatment. The equation is assumed to hold for all subjects in a study.

NMAR is a particular concern with weight loss studies because low values for R may be more likely to occur for a subject who is not losing weight. In such settings, the missing values for that subject may be NMAR unless the subject stayed in the trial long enough for data to be collected describing the subject's 'lack of weight loss'. If the subject's lack of weight loss is observed with existing data, then the data are said to be MAR. A subject's missing values are only MCAR if the reason for dropout (i.e. the value of R) was unrelated to the weight loss for that subject (e.g. a subject's employment requires moving out of the area). In any given study, the distinction regarding whether the missing data are MAR, MCAR, or NMAR is not a trivial one, because the choice of analysis method is highly dependent upon the type of missing data mechanism.

Table 1 A monotone missingness pattern for four subjects with different values for the dropout variable, R

Subject	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	R
1	$y_{1,0}$	$y_{1,1}$	$y_{1,2}$	$y_{1,3}$	?	?	3
2	$y_{2,0}$	$y_{2,1}$	$y_{2,2}$	?	?	?	2
3	$y_{3,0}$	?	?	?	?	?	0
4	$y_{4,0}$	$y_{4,1}$	$y_{4,2}$	$y_{4,3}$	$y_{4,4}$	$y_{4,5}$	5

Y_0 indicates a baseline measurement and Yj, the change from baseline in the outcome variable at the jth follow-up visit. The lower case $y_{i,j}$ indicates an observed value for the ith subject at the jth follow-up visit, and missing observations are denoted with a '?'.

Table 2 Dropout mechanism defined as a probability statement, and the resulting type of missing data

Definition of dropout mechanism	Notation for missing data	Ignorability of missing data mechanism?
$P(R = r\|Y_{obs}, Y_{mis}, X) = P(R = r)$	MCAR	Yes
$P(R = r\|Y_{obs}, Y_{mis}, X) = P(R = r\|X)$	MCAR2	Yes, if Y_{obs} and X used in method of analysis
$P(R = r\|Y_{obs}, Y_{mis}, X) = P(R = r\|Y_{obs}, X)$	MAR	Yes, if and X used in method of analysis
$P(R = r\|Y_{obs}, Y_{mis}, X) = P(R = r\|Y_{obs}, Y_{mis}, X)$	NMAR	No, the relationship between R and Y_{mis} cannot be eliminated using available information

Observed values for the outcome variable (e.g. change from baseline) are Y_{obs}, missing values because of dropout are Y_{mis}, and other measured covariates are denoted by X. A monotone dropout mechanism is assumed and R is equal to the number of the last follow-up visit before the subject dropped out.

In Table 1 for subject 1, $R = 3$ and we observe a value for Y_3 but the design requires observing a value for Y_5. Because Y_5 is unobservable, the subject may either be omitted from the analysis (as in a complete case analysis), or Y_5 may be estimated (imputed). Later, we discuss a method involving mixed effects models where the subject is neither omitted from the analysis nor is their missing response imputed.

Imputation strategies and two issues

There are two issues to consider when imputing missing data. First, using a technique that assumes missing data are MCAR (or MCAR2) when they are actually MAR can bias conclusions from a study just as can assuming missing data are MAR when they are actually NMAR. Thus MAR is seen as a less restrictive assumption than MCAR or MCAR2 and may often be more reasonable in practice (35). A second issue involves the underestimation of the variance of an estimated treatment effect.

One *ad hoc* method that has been used by researchers in the past is to fill in missing values with the mean for all non-missing subjects for that value. For example, if a covariate of weight at 3 months was missing for the *i*th subject, we would compute the mean weight at 3 months for all non-missing subjects and *impute* this value for the *i*th subject. LOCF is another method that imputes a value for a subject who has dropped out of a study. LOCF and the method of replacing missing data with mean values requires the missing values to be MCAR (47) (or MCAR2 if X is used in the analysis) to yield unbiased estimates of a treatment effect. Moreover, both of these methods also underestimate standard errors resulting in inflation of the true probability of observing a type 1 error in the study. More sophisticated techniques for imputing missing data that only require MAR have been proposed to address the first issue (cf. 38), and a method of 'multiple' imputation has been proposed to address the second (48).

Under MAR, one can obtain unbiased estimates of the treatment effect by using more complicated models for replacing the missing values. If the observed variables thought to be related to the missing data are included in the model used to impute the missing values, then this approach can lead to unbiased parameter estimates. For example, if it is thought that men are more likely than women to drop out of the study, one can replace the missing values with the gender-specific means. Obviously, this approach can be generalized to complex models for describing the missing values in which a large number of covariates may be involved. The advantage of such an approach is that, after values for the missing data have been imputed, one can use standard methods of analysis that would have been used had complete data been obtained from the start.

The problem related to underestimating the standard errors of the estimated treatment effect does not go away even when more complex models are used for the imputation (46). This results from the fact that we are, in essence, pretending that we have complete data during the analysis phase and are ignoring the variability inherent in the imputation process. Thus, instead of imputing a single value, one should use multiple imputation and replace each missing value with a set of plausible random values in order to capture the variability surrounding the missing value (48). If the imputation process is repeated k times, a set of k multiply imputed data sets would be created that can each be analysed using a statistical procedure applicable for complete data. As a consequence, one is left with k sets of parameter estimates. Final estimates are obtained by combining the k sets of parameter estimates. The variability of the final estimates incorporates the inherent variability of the imputation process. It has been shown that this process avoids the underestimation of standard errors inherent with single imputation with as few as 5–10 imputations (32).

Until recently, researchers had to use special software or write their own programs to implement multiple imputation. However, with version 8.1, the SAS Institute introduced two new procedures that enable one to implement multiple imputation: PROC MI and PROC MIANALYSE (49). Multiple imputation using SAS proceeds in three distinct phases:

(1) Use PROC MI to create the k multiply imputed data sets.

(2) Use some standard SAS procedure to analyse these k data sets and save the parameter estimates and standard errors into a data set. This involves the same procedures that would be used to analyse a complete data set with no missing data.

(3) Use PROC MIANALYSE to combine the k estimated parameters and standard errors into the final estimates.

PROC MI uses three distinct methods for imputation: (1) a parametric regression method that assumes multivariate normality; (2) a non-parametric method that uses propensity scores; and (3) a Markov chain Monte Carlo (MCMC) method that assumes multivariate normality. The first two methods are useful with monotone missing data patterns. The MCMC method is recommended when the pattern of missing data is arbitrary.

Multiple imputation can be performed in other software packages as well. Horton & Lipsitz (50) provide an excellent comparison of several such software packages. We mention two. SPSS is a package familiar to many obesity researchers. The Missing Values Analysis (MVA) module in SPSS allows one to implement multiple imputation to create the k multiply imputed data sets (51). The user can then use standard SPSS procedures to analyse these k data sets. However, the user must then combine the parameter esti-

mates from these k data sets in the appropriate manner. The MVA module in SPSS is analogous to PROC MI in SAS but there is no SPSS analogue to the PROC MIANALYSE procedure in SAS. There are also multiple imputation and analysis software routines written for the statistical package S-Plus (52) that are described, and an internet site for them given, in Schafer (46,53).

Because researchers must specify a model for imputation, and the validity of this model affects the validity of the conclusions, one should give the imputation model as much thought as the analysis model during the design phase of a study. It is not necessary that the same set of covariates are included in both models, and it is often a good idea to include a large number of covariates in the imputation model because the precision lost by including any predictors not useful for estimating the missing value is much less than the price of incorrectly specifying the model. The imputation model can include enough pre-withdrawal information to make the result of assuming MAR plausible in practice (32). Multiple imputations created under an incorrect model can lead to incorrect conclusions (54).

Hunsberger *et al.* (35) describe their experience in choosing an imputation strategy for a multicentre school-based study to compare percent body fat between treatment and usual care groups at the end of a 3-year obesity prevention intervention in American Indian children. They found that a multiple imputation approach using a regression equation based on subjects with observed data had the best performance with regards to type 1 error and power. Allison *et al.* (27) used a multiple imputation procedure as their primary analysis technique of data from an obesity trial, and they supplemented these results with those obtained from a complete case analysis. Although results from the two were not expected to be the same, the secondary analysis was considered an added step to highlight any major differences that might require more detailed investigation.

Mixed effects models

A statistical method that does not require imputing missing values has been developed that is based on maximum likelihood inference (55). This method often goes by several names including mixed linear models, two-stage random effects models, or random coefficient models (39). For longitudinal studies where missing data may be observed over the course of the study, a repeated measures mixed model requires only that missing data be MAR, and it may be more robust to potential bias from missing data than the traditional *ad hoc* methods (LOCF, completers only) when the data are not MCAR (47). As opposed to traditional repeated measurement techniques, mixed models permit the inclusion of patients with missing values at some time points. Often, the primary endpoint may be the change from baseline to some pre-specified time points. Mixed models are preferable to *ad hoc* methods in these situations. However, the greatest utility of such mixed models is the ability to model trends over time. For example, two groups may be similar in their outcome values at the end of the study but may differ drastically in how they arrive at those values. Such analyses are difficult to perform using an LOCF model because the very nature of carrying the last observation forward diminishes the perceived variability in the data over time.

Mixed models may be used to evaluate linear, quadratic, or higher order trends over time separately for each group of interest, while simultaneously controlling for any covariates of interest. In such studies, the measurements observed over time within a patient are correlated, and it is often the case that the variation in measurements increase over time. In order to account for both the correlations over time within a patient and the increasing variation from visit to visit, random intercepts and slopes may be fit to the observations from any specific subject. As a consequence, we allow the model to differ for each person, hence the name 'random effects' model. By taking into account the correlation of measurements over time for any particular subject, we can obtain more efficient estimates of the treatment effect at any particular time point. There are several texts now available discussing the use of mixed effects models (e.g. 56,57), and other texts specifically address their implementation in S-Plus (58) and in SAS (59).

The issue of data NMAR

When the data are thought to be NMAR, the options for analysis are more limited. Although some options have been suggested in the statistical literature (38,44,60,61), they are less easy to implement with standard software. For that reason, we do not discuss details of these methods here.

There is no magic solution to dealing with data that are NMAR and care must be taken when drawing conclusions based upon how some subjects 'might' have responded had they not dropped out (62). Overall *et al.* (63) highlighted this using simulations to evaluate type 1 error and power for different methods of data analysis and different dropout mechanisms. It is recommended that a statistician be involved in the design of, and the analysis of data from, an obesity trial, particularly if missing data that may be NMAR are expected.

A simulation example

A small simulation was run to compare methods of analysis using complete cases only, LOCF, multiple imputation, and mixed effect models when data are MCAR, MAR, and NMAR. More elaborate simulations have been reported elsewhere (32,47,61). We suppose that 100 subjects are

randomized into two treatment groups of 50 subjects each. Weight in kilograms (kg) is measured at baseline (visit 1) and three follow-up visits at 4-week intervals, that is, 4, 8, and 12 weeks. The parameter of interest is mean difference in weight loss between the two treatment groups after 12 weeks.

The complete data were generated similarly to Liu & Gould (32). Data were multivariate normal with mean vector for the treatment group equal to $\mu_T = [92,88,85,82]$, and for the second (control) group equal to $\mu_C = [92,90,90,89]$. The two mean vectors were arbitrarily selected as was done in Liu & Gould (32) and here they reflect a 12-week 'treatment' effect of 7 kg additional weight loss in the treatment group (positive outcomes denote weight 'loss'). Standard deviations were selected using available cases in data used in Allison *et al.* (27) and were equal to [14.3,14.0,14.2,13.9]. The correlation structure was also modelled after the data in Allison *et al.* (27) and is given by the four-dimensional correlation matrix:

$$\begin{bmatrix} 1 & 0.986 & 0.967 & 0.949 \\ 0.986 & 1 & 0.992 & 0.980 \\ 0.967 & 0.992 & 1 & 0.995 \\ 0.949 & 0.980 & 0.995 & 1 \end{bmatrix}.$$

Thus, there is very high correlation between outcomes at each subsequent visit with higher correlations occurring at time points separated by only 4 weeks.

All 'subjects' in the simulation returned for the first visit and then either remained or dropped out according to a monotone pattern. Let *Cj* indicate a change from baseline at visit *j* (Baseline–Visit *j* so positive values are weight 'loss', $j = 2, 3, 4$). To simulate missing data that are MCAR or MAR, subjects 'missed' the final 12-week visit with probability p_{14} if $C_3 \leq 0$ and with probability p_{24} if $C_3 > 0$. Subjects missed both the 8- and 12-week visits with probability p_{13} if $C_2 \leq 0$ and with probability p_{23} if $C_2 > 0$. Missing data that were MCAR were generated using $p_{14} = p_{24} = 0.3$ and $p_{13} = p_{23} = 0.15$. Missing data that were MAR were generated using $p_{14} = 0.6$, $p_{24} = 0.2$ and $p_{13} = 0.4$, $p_{23} = 0.1$. In the latter case, subjects who did not lose weight by the 4-week visit were more likely to drop out of the study. Similarly, subjects who showed up for the 8-week visit and had not lost weight were more likely to miss the 12-week visit. In the MCAR situation, the probabilities were the same.

NMAR data were created by slightly altering the definition of the above probabilities. Subjects missed the final 12-week visit with probability p_{14} if $C_4 \leq 0$ and with probability p_{24} if $C_4 > 0$. Subjects missed both the 8- and 12-week visits with probability p_{13} if $C_3 \leq 0$ and with probability p_{23} if $C_3 > 0$. The actual values for p_{14}, p_{24}, p_{13}, p_{23} were the same as those for the MAR situation. So a subject dropped out of the trial with a probability depending on what their weight change *would have been* had they showed up, a probability depending on a missing value. The numbers of missing observations at 12 weeks were similar for the three cases and varied over simulations from 20 to 50%.

A complete case analysis for each data set was conducted using the usual two-sample pooled variance *t*-tests on weight change at 12 weeks for the subjects that completed the study. LOCF analysis was conducted using the same two-sample *t*-tests on the completed data set of 100 observations. Multiple imputation was conducted using a regression imputation procedure described in detail in Liu & Gould (32). Five imputations were used with each data set, the same *t*-tests conducted on each completed data set, and the results combined as described in Liu & Gould (32). A mixed effect model analysis was conducted by coding time as a factor variable with levels for 4, 8, and 12 weeks, and the treatment indicator variable was equal to 1 for treatment group and 0 for control. The response variable was change from baseline at each subsequent visit. The chosen model included treatment, time, and a treatment by time interaction as explanatory fixed effects, and a random effect for subject and time. A predicted 'treatment effect' was estimated from the model at the 12th week.

For each analysis an estimate of the mean treatment effect, its standard error, and a 95% confidence interval was recorded for each of 1000 simulations. In the case of the mixed model analysis, confidence intervals were approximate but expected to be fairly accurate because of the reasonably large sample size [see Chapter 22 of Milliken & Johnson (64) for more detail]. Results from the simulation are shown in Table 3 and discussed in the next

Table 3 Results from 1000 simulations

	MCAR	MAR	NMAR
Complete cases	6.96, 1.12, 0.956	6.09, 1.05, 0.862	6.26, 1.09, 0.895
LOCF	5.98, 0.88, 0.791	6.18, 0.89, 0.853	6.23, 0.88, 0.852
Multiple imputation	6.99, 0.95, 0.958	6.99, 0.95, 0.961	7.00, 0.94, 0.961
Mixed effects models	7.08, 0.94, 0.948	7.09, 0.93, 0.951	7.10, 0.93, 0.953

The three numbers shown in each of the 12 entries are the mean of the 1000 estimated treatment effects, the mean of 1000 standard errors, and the proportion coverage of 95% confidence intervals, respectively, for the three missing data patterns and four methods of analysis. The true mean effect = 7 kg; MCAR, missing completely at random; MAR, missing at random; NMAR, not missing at random; LOCF, last observation carried forward.

section. The simulations were coded and executed in S-Plus; however, we include some SAS code to analyse a sample data set in the Appendix. This data set would be called TESTDATA and includes variables Y (change from baseline), TRT (the treatment variable), Time (time of follow-up visit), BASE (baseline weight), and SUB (subject number).

Results and summary

The results in Table 3 illustrate many of the ideas discussed earlier. The complete case analysis generally obtains an unbiased estimate of the mean treatment effect when missing data are MCAR but produces biased estimates under MAR or NMAR. Standard errors are higher than the other methods of analysis because of loss of efficiency associated with smaller sample sizes (after omitting cases with missing values). LOCF produces biased estimates for all three types of missing data in these particular simulations. This happens because LOCF misses the trend of 'more' weight loss as subjects remain in the trial longer. The LOCF estimates are conservative in these simulations. LOCF also produces the smallest standard errors because values imputed via the LOCF method are taken as fixed at the final visit. With the exception of complete case analysis for MCAR data, confidence interval coverage is not accurate for either complete case analysis or LOCF because of biased estimates and, in the case of LOCF, underestimation of standard errors.

Both multiple imputation and mixed effects models appear to produce unbiased estimates of a treatment effect for all types of missing data. Moreover, standard errors are fairly consistent between the two methods and among the types of missing data. Confidence interval coverage is also accurate with multiple imputation being perhaps slightly conservative. Most noteworthy is that both multiple imputation and mixed effects models perform well for either data MAR or NMAR. The fact that all subjects completed the first follow-up visit (and that there were only three scheduled) combined with the high correlation of weight in the multivariate normal correlation matrix contributed to this robust behaviour for NMAR data. Multiple imputation and mixed effects models were able to capture enough information in the earlier visits to produce valid estimates of a treatment effect.

Which of the two methods, that is, multiple imputation or mixed effects models, is superior in general remains open to question. Both have been used to advantage for the analysis of incomplete data sets. The mixed effects models might have a slight edge in preference for various reasons. Some issues have been brought up regarding multiple imputation when the analyst's and imputer's models are vastly different. This was thought possible with large survey data sets (perhaps from government sources) where the imputer and analyst may be different individuals in different organizations. Whether such situations could commonly occur with clinical trial data is less certain.

The mixed effects models have now been supplemented with routines in statistical analysis software and supporting textbooks dealing specifically with implementation of the models using that software. Their connection to traditional linear regression models may also have an appeal in addition to their mathematical foundation, that is, maximum likelihood inference. They are not, however, impervious to producing errors. Other simulations of the mixed effects model analysis (not shown here) illustrated biased estimates, underestimation of variances, and inaccurate confidence interval coverage. These situations occurred when the model was misspecified. For example, coding time as a numeric variable and including it in the model as a linear term overestimates the treatment effect. Observing that the population mean treatment effects at weeks 4, 8, and 12 are 2, 5, and 7 kg, respectively, including only a linear term with time in the model will establish a trend that overestimates the week 12 weight change. Moreover, if time is not included as a random effect, the high (yet varying) correlation from visit to visit underestimates standard errors. This highlights the need for the same tests of assumptions and diagnostic techniques as required for usual linear regression. Misspecifying a model can lead to errors regardless of the type of missing data. A summary of the information on missing data, recommended methods of analysis, and key references are given in Table 4.

Table 4 Summary of 'modern' methods of analysis, appropriate use and limitations, and some selected applied and theoretical references

Method of analysis	Appropriate use	Applied references	Theory references
Multiple imputation using means of observed data	When missing data are MCAR	34	38, 42, 45, 48
Multiple imputation with regression models using X	When missing data are MCAR2	34	38, 42, 45, 48
Multiple imputation with models using all Y_{obs}, X	When missing data are MAR	27, 32, 34, 35, **49, 50, 51, 53**	38, 42, 45, 48, 54
Mixed effect models, assumed to incorporate Y_{obs}, X	When missing data are MAR	33, 34, 47, **58, 59,** 63	39, 55, 56, 57
Pattern mixture models	Potentially useful to detect consequences of NMAR	61	38, 43, 60

References highlighted in bold indicate those focused on a software implementation of the method. MCAR, missing completely at random; MAR, missing at random; NMAR, not missing at random.

Acknowledgements

This research supported in part by NIH grant P30DK056336. The authors thank two anonymous referees for helpful suggestions that improved the manuscript.

References

1. Douketis JD. The long-term effectiveness of weight reduction interventions in patients with obesity: a critical review of the literature. *JCOM* 2000; **7**: 31–44.
2. Mokad AH, Ford ES, Bowman BA, Dietz WH, Vinicor F, Bales VS, Marks JS. Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. *JAMA* 2003; **289**: 76–79.
3. Kolotkin RL, Meter K, Williams GR. Quality of life and obesity. *Obes Rev* 2001; **2**: 219–229.
4. Fontaine KR, Cheskin LJ, Barofsky I. Health related quality of life in obese persons seeking treatment. *J Fam Pract* 1996; **43**: 265–270.
5. Fontaine KR, Barofsky I, Cheskin LJ. Predictors of quality of life for obese persons. *J Nerv Men Dis* 1997; **185**: 120–122.
6. Heo M, Allison DB, Faith MS, Zhu S, Fontaine KR. Obesity and health-related quality of life: the potential mediating effects of joint pain and comorbitities. *Obes Res* 2003; **11**: 209–216.
7. Colditz GA. Economic costs of obesity. *Am J Clin Nutr* 1992; **55**: 5035–5075.
8. Wolf AM, Coditz GA. Social and economic effects of body weight in the United States. *Am J Clin Nutr* 1996; **63**: S466– S469.
9. Mitka M. Economist takes aim at 'big fat' US lifestyle. *JAMA* 2003; **289**: 33–34.
10. Allison DB, Zannolli R, Narayan KM. The direct health care costs of obesity in the United States. *Am J Public Health* 1999; **89**: 1194–1199.
11. Allison DB, Zannolli R, Faith MS, Heo M, Petrobelli A, Van Itallie TB, Pi-Sunyer FX, Heymsfield SB. Weight loss increases and fat loss decreases all-cause mortality rate: results from two independent cohort studies. *Int J Obes* 1999; **23**: 603–611.
12. Allison DB, Fontaine KR, Manson J, Stevens J, Van Itallie TB. How many deaths are attributable to obesity in the United States. *JAMA* 1999; **282**: 1530–1538.
13. Fontaine KR, Redden DT, Wang C, Westfall AO, Allison DB. Years of life lost due to obesity. *JAMA* 2003; **289**: 187–193.
14. Sorensen TIA. Weight loss causes increased mortality: pros. *Obes Rev* 2003; **4**: 3–7.
15. Yang D, Fontaine KR, Wang C, Allison DB. Weight loss causes increased mortality: cons. *Obes Rev* 2003; **4**: 9–16.
16. Popkin B, Doak C. The obesity epidemic is a worldwide phenomenon. *Nutr Rev* 1998; **56**: 106–114.
17. Ebbeling CB, Pawlak DB, Ludwig DS. Childhood obesity: public-health crisis, common sense cure. *Lancet* 2002; **360**: 473–482.
18. Gunnell DJ, Frankel SJ, Kiran N, Peters TJ, Smith GD. Childhood obesity and adult cardiovascular mortality: a 57-y follow-up study based on the Boyd Orr cohort. *Am J Clin Nutr* 1998; **67**: 1111–1118.
19. Allison DB, Kaprio J, Korkeila M, Koskenvuo M, Neale MC, Hayakawa K. The heritability of BMI among an international sample of monozygotic twins reared apart. *Int J Obes* 1996; **20**: 501–506.
20. Faith MS, Tepper BJ, Hoffman DJ, Pietrobelli A. Genetic and environmental influences on childhood obesity. *Clin Fam Pract* (in press).
21. Segal N, Allison DB. Twins and virtual twins: bases of relative body weight revisited. *Int J Obes* 2002; **4**: 437–441.
22. Crawford D. Population Strategies to prevent obesity. *BMJ* 2002; **325**: 728–729.
23. National Institutes of Health. *The Practical Guide: Identification, Evaluation, and Treatment of Overweight and Obesity in Adults*. Publication number 00-4084. The National Institutes of Health: Bethesda, 2000.
24. Pi-Sunyer X. A clinical view of the obesity problem. *Science* 2003; **299**: 859–860.
25. Astrup A. Healthy lifestyles in Europe: prevention of obesity and type II diabetes by diet and physical activity. *Public Health Nutr* 2001; **4**: 499–515.
26. Astrup A, Rossner S. Lessons from obesity management programmes: greater initial weight loss improves long-term maintenance. *Obes Rev* 2000; **1**: 17–19.
27. Allison DB, Gadbury G, Schwartz LG, Murugesan R, Kraker JL, Heshka S, Fontaine KR, Heymsfield SB. A novel soy-based meal replacement formula for weight loss among obese individuals: a randomized controlled clinical trial. *Eur J Clin Nutr* 2003; **57**: 514–522.
28. Gray-Donald K, Kramer MS. Causality inference in observational vs. experimental studies, an empirical comparison. *Am J Epidemiol* 1988; **124**: 885–892.
29. Food and Drug Administration. *Guidance for the Clinical Evaluation of Weight-Control Drugs*. Division of Metabolic and Endocrine Drug Products, Food and Drug Administration: Rockville, 24 September 1996.
30. Smith IG, Goulder MA. Randomized placebo-controlled trial of long term treatment with sibutramine in mild to moderate obesity. *J Fam Pract* 2001; **50**: 505–512.
31. Jain AK, Kaplan RA, Gadde KM, Wadden TA, Allison DB, Brewer ER, Leadbetter RA, Richard N, Haight B, Jamerson BD, Buaron KS, Metz A. Buropion SR vs. placebo for weight loss in obese patients with depressive symptoms. *Obes Res* 2002; **10**: 1049–1056.
32. Liu G, Gould AL. Comparison of alternative strategies for analysis of longitudinal trials with dropouts. *J Biopharm Stat* 2002; **12**: 207–226.
33. Heyting A, Tolboom JTBM, Essers JGA. Statistical handling of drop-outs in longitudinal clinical trials. *Stat Med* 1992; **11**: 2043–2061.
34. Albert PS. Tutorial in Biostatistics: longitudinal data analysis (repeated measures) in clinical trials. *Stat Med* 1999; **18**: 1707–1732.
35. Hunsberger S, Murray D, Davis CE, Fabsitz RR. Imputation strategies for missing data in a school-based multi-center study: the Pathways study. *Stat Med* 2001; **20**: 305–316.
36. Streiner DL. The case of the missing data: methods of dealing with dropouts and other research vagaries. *Can J Psychiatry* 2002; **47**: 68–75.
37. Mazumdar S, Liu KS, Houck PR, Reynolds CF. Intent-to-treat analysis for longitudinal clinical trials: coping with the challenge of missing values. *J Psychiatric Res* 1999; **33**: 87–95.
38. Little RJA, Rubin DB. *Statistical Analysis of Missing Data*. John Wiley & Sons: New York, 1987.
39. Cnaan A, Laird NM, Slasor P. Tutorial in biostatistics: using the general linear mixed model to analyze unbalanced repeated measures longitudinal data. *Stat Med* 1997; **16**: 2349–2380.
40. Greenway FL, Bray GA, Marlin RL. Methods to maximize retention in weight loss studies. *Obes Res* 1999; **7**: 593–596.
41. Allison DB, Cappelleri JC, Carpenter KM. Design and analysis of obesity treatment and prevention trials. In: Dalton S (ed.)

Overweight and Weight Management. ASPEN Publications: Gaithersburg, MD, 1997, pp 557–595.
42. Allison DB, Pi-Sunyer FX. Obesity treatment: examining the premises. *Endocr Pract* 1995; **1**: 353–364.
43. Rubin DB. Inference and missing data. *Biometrika* 1976; **63**: 581–592.
44. Little RJA. Modeling the drop-out mechanism in repeated-measures studies. *J Am Statist Assoc* 1995; **90**: 1112–1121.
45. Heitjan DF. Ignorability and bias in clinical trials. *Stat Med* 1999; **18**: 2421–2434.
46. Shafer JL. *Analysis of Incomplete Multivariate Data*. Chapman & Hall: New York, 1997.
47. Mallinckrodt CH, Clark WS, David SR. Accounting for drop-out bias using mixed-effects models. *J Biopharm Stat* 2001; **11**: 9–21.
48. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons: New York, 1987.
49. SAS Institute Inc. *SAS/STAT User's Guide, Version 8*. SAS Institute, Inc.: Cary, NC, 2001.
50. Horton NJ, Lipsitz SR. Multiple imputation in practice: comparison of software packages for regression models with missing variables. *Am Statistician* 2001; **55**: 244–254.
51. SPSS Inc. *SPSS Missing Value Analysis 7.5*. MaryAnn Hill/SPSS Inc.: Chicago, IL, 1997.
52. Insightful Corporation. *S-Plus 6 for Windows User's Guide*. Insightful Corp.: Seattle, WA, 2001.
53. Schafer JL. (1997) *Software for multiple imputation*. [WWW internet site]. URL: http://www.stat.psu.edu/~jls/
54. Fay RE. Alternative paradigms for the analysis of imputed survey data. *J Am Statist Assoc* 1995; **91**: 490–498.
55. Laird NM, Ware JH. Random effects models for longitudinal data: an overview of recent results. *Biometrics* 1982; **38**: 963–974.
56. Longford NT. *Random Coefficient Models*. Clarendon Press: Oxford, 1993.
57. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. Springer: New York, 2000.
58. Pinheiro JC, Bates DM. *Mixed-Effects Models in S and S-Plus*. Springer: New York, 2000.
59. Littell RC, Milliken GA, Stroup WW, Wolfinger RD. *SAS System for Mixed Models*. SAS Institute, Inc.: Cary, NC, 1996.
60. Wei L, Shih WJ. Partial imputation approach to analysis of repeated measurements with dependent drop-outs. *Stat Med* 2001; **20**: 1197–1214.
61. 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. *J Biopharm Stat* 1998; **8**: 545–563.
62. Rubin DB. Formalizing subjective notions about the effect of nonrespondents in sample surveys. *J Am Statist Assoc* 1977; **72**: 538–543.
63. Overall JE, Shobaki G, Fiore J. Random regression with imputed values for dropouts. *Psychopharmacol Bull* 1996; **32**: 377–388.
64. Milliken GA, Johnson DE. *Analysis of Messy Data, Volume 1: Designed Experiments*. Wadsworth: Belmont, CA, 1984.

Appendix

SAS code for analysing missing data example

```
/* Sort data */

  PROC SORT DATA=TESTDATA;
    BY SUB TIME;
  RUN;

  /* Complete Case Analysis */

  PROC TTEST DATA=TESTDATA;
    WHERE TIME=12;
    CLASS TRT;
    VAR Y;
  RUN;

  /* LOCF Analysis */

  DATA LOCF;
    SET TESTDATA;
    BY SUB TIME;
    RETAIN LOCFY;
    IF FIRST.SUB THEN LOCFY=.;
    IF Y NE . THEN LOCFY=Y;
  RUN;

  PROC TTEST DATA=LOCF;
    WHERE TIME=12;
    CLASS TRT;
    VAR LOCFY;
  RUN;

  /* Multiple Imputation Analysis */

  PROC TRANSPOSE DATA=TESTDATA OUT=TEMP;
    BY SUB;
    VAR Y;
    ID TIME;
    COPY BASE TRT;
  RUN;

  DATA TESTDATAB;
    SET TEMP;
    WHERE _NAME_='Y';
  RUN;

  PROC MI DATA=TESTDATAB SEED=345621 OUT=
OUTMI;
    MONOTONE METHOD=REG;
    VAR _4 _8 _12;
  RUN;

  PROC SORT DATA=OUTMI;
    BY _IMPUTATION_;
  RUN;

  PROC GLM DATA=OUTMI;
    BY _IMPUTATION_;
    MODEL _12 = TRT / SOLUTION INVERSE;
    ODS OUTPUT PARAMETERESTIMATES=GLMPARMS
               INVXPX=GLMXPXI;
  RUN;

  PROC MIANALYSE PARMS=GLMPARMS XPXI=
GLMXPXI;
```

```
  VAR INTERCEPT TRT;
RUN;

/* Mixed Model Analysis */

PROC MIXED DATA=TESTDATA;
  CLASS TRT TIME;
    MODEL Y = TRT TIME TRT*TIME /
 SOLUTION;
    RANDOM INTERCEPT TIME / SUBJECT=SUB;
    ESTIMATE '12 week Effect' TRT 1
-1 TRT*TIME 0 0 1 0 0 -1 / CL;
 RUN;
```