
A cautionary tale: dealing with missing data in clinical trials for rheumatic diseases*

J. Song, W.J. Boscardin, D.E. Furst, D. Khanna, W.K. Wong

*This is an adaptation from a project for the Scleroderma Foundation (SF) in which statistical software for analysing missing data was made available for clinicians through the SF.

Juwon Song, PhD
W. John Boscardin, PhD
Daniel E. Furst, MD
Dinesh Khanna, MD, MS
Weng Kee Wong, PhD

Please address correspondence to:

Daniel E. Furst, MD.
UCLA Medical School,
1000 Veteran Ave, Rm 32-59,
90025 Los Angeles, CA, USA.
E-mail: defurst@mednet.ucla.edu

Received on February 20, 2014; accepted in revised form on September 18, 2014.

Clin Exp Rheumatol 2014; 32 (Suppl. 86): S122-S126.

© Copyright CLINICAL AND EXPERIMENTAL RHEUMATOLOGY 2014.

Key words: missing data, ignorable missing data, non-ignorable missing data, SSc clinical trials

Competing interests:

D. Furst has received grant or research support from AbbVie, Actelion, Amgen, BMS, Gilead, GSK, NIH, Novartis, Pfizer, Roche/Genentech, and UCB. He serves as a consultant for AbbVie, Actelion, Amgen, BMS, Janssen, Gilead, GSK, NIH, Novartis, Pfizer, Genentech/Roche, and UCB, and also serves on the speaker bureau for AbbVie, Actelion, and UCB (for CME only). D. Khanna reports grants from NIAMS (K24) 5-K24-AR-063120-03, NIAID and personal fees from Actelion, Bayer, DIGNA, Genentech/Roche, Gilead, Lycera, Merck, Takeda, and United Therapeutics. Grants and personal fees from BMS, Savient, InterMune outside the submitted work. The other co-authors have declared no competing interests.

ABSTRACT

Objective. Missing data are found in nearly all clinical trials and it is important to use appropriate statistical techniques to analyse clinical trials with missing data. We discuss common statistical methods for tackling missing data and how to handle results when the analyses give different results.

Methods. Using data from a placebo-controlled, randomised bovine Type I collagen (CI) study in diffuse cutaneous systemic sclerosis (dcSSc), we apply different statistical approaches to handling missing data. We also describe simple ways to ascertain the type of missing data in the data set, to the extent possible.

Results. We examine eleven different methods to impute missing data. An analysis based on completers alone (complete case analysis and available case analysis) and the last observation carried forward (LOCF) methods require underlying assumptions which are rarely met in practice. Multiple imputation, mixed effects, and repeated measures try to account for the differences among patients and account for patient's specific response patterns, although the assumption that the missing data is directly related to the observed characteristics may well not be true.

The joint likelihood based model combines the mixed effect model and logistic regression model to explicitly handle data not missing at random and so it is more realistic and potentially takes an additional step toward decreasing bias.

Conclusion. We discussed various ways of handling missing data and provide recommendations on how to arrive at a conclusion when different statistical approaches to analyse missing data analysis in clinical trials give conflicting answers.

Introduction

Statistical methods for analysing missing data have advanced a great deal in the last several decades. Despite this, data from many clinical trials in rheumatic

diseases continue to be analysed using methodology that may be inappropriate for handling missing data. Frequently, an intention to treat (ITT) analysis with last observation carried forward (LOCF) is used for analysing missing longitudinal outcome data. However, this approach carries “inherent risks and in most cases is unjustifiable” (Carpenter *et al.*, 2004). We will review and explain methods that are more robust and less biased for handling missing longitudinal outcome data. Using data from a double-blind, randomised, placebo-controlled trial of bovine collagen (Postelthwaite *et al.*, 2008), we show the effect of various methods on the resulting conclusions and how one may reconcile conflicting results.

Methodology

Definitions

Missing data may occur completely at random, missing at random, or not missing at random (Rubin, 1976). Data is said to be missing completely at random (MCAR) if the probability of missingness does not depend on any observed or even unobserved data. This might hold, for example, when the physician does not come to the patient visit for reasons unrelated to the patient, or the instrument to obtain laboratory data malfunctions. Missing data is said to occur at random (MAR) if it may depend on observed data but not on unobserved data. For example if a patient is more likely to miss a visit if her condition at the previous measured visit had improved; then these data are said to be MAR (but not MCAR, since the probability of missingness depended on observed data). Data are missing not at random (NMAR) if the probability of missingness depends on unobserved data. For example, a clinic-based measurement such as HAQ-DI, might be missing because the patient has such severe functional disability that she cannot leave home for the visit. Further discussion and applications of these concepts in the

analysis of missing data in clinical trials are available in Molenberghs and Kenward (2007).

Data setting

We use the data from the double-blind, randomised, placebo-controlled trial of bovine CI in dcSSc to illustrate these concepts and perform the analyses based on different assumptions on the nature of the missing data.

Briefly, in this trial, 168 eligible patients with dcSSc were enrolled in a multicentre, double-blind, placebo-controlled trial of bovine CI *versus* placebo. Patients were predefined into *early disease* (disease duration ≤ 3 years) and *late disease* (disease duration 4 to 10 years). The total duration of the treatment phase was 12 months with a follow-up visit at month 15 (three months off study medication for safety follow-up). The principal outcome in this study was the modified Rodnan skin score (MRSS), comparing its change in the bovine CI group to the placebo group at 12 months (Postlethwaite *et al.*, 2008). As in most trials, there were dropouts and missed visits resulting in missing data.

Missingness description and considerations

Missed visits due to chance circumstances unrelated to the patient's condition would result in MCAR data. Many longitudinal studies have MCAR data, which is probably too strong an assumption for practical applications. Generally, missing data could be considered as either MAR or NMAR. The assumption of MCAR may be verified by comparing fully observed characteristics between compliers and non-compliers (Little, 1988). However, it is not possible to verify whether the missing data mechanism is MAR or NMAR based on a given incomplete data set.

Analysis methods

General methods for analysing longitudinal clinical trial data with missing data are shown in Table I, along with their principal positive and negative attributes. We analyse the bovine CI data using eleven specific methods including two relatively simple methods (complete case analysis and available case analysis), a single imputation method (last

observation carried forward), two likelihood methods assuming MAR (mixed effects and repeated measures), and two likelihood methods assuming NMAR (a joint mixed effects plus missing data modelling for a single longitudinal outcome (Carpenter *et al.* 2002). We demonstrate one additional method which is an extension of this *double* model so that multivariate longitudinal outcomes can be considered (Boscardin *et al.*, 2007). Using a fully conditional specification, we multiply imputed the bovine CI data five times using the MICE package in R. We then analyse the multiply imputed bovine CI data using four methods including complete case analysis, mixed effects modelling, repeated measures modelling and generalised estimating equations (GEE). Software to fit these models along with details on the specific assumptions used for the eleven analyses presented here can be obtained by emailing the first author.

Results

The main published results of this study showed that bovine CI did not change the mRSS more than placebo at 12 and 15 months in the total patient population. However, in sub-analysis the *late disease* group (>3 –10 years of disease) treated with bovine CI improved more than placebo at 15 months although not at 12 months. This suggested that the bovine CI may benefit late phase SSC patients in a delayed manner (Postlethwaite *et al.*, 2008).

The above results were arrived at after analysing missing data in multiple ways, where different analyses gave different estimates of responses at 12 and 15 months. There are assumptions required for each method to be valid, including assumptions on the type of the missingness in the data. It is therefore important to try to diagnose the type of missingness we have in the data as a first step in analysing missing data. Graphically, it is possible to investigate whether missingness is plausibly MCAR. Figure 1 shows completers in the bovine CI trial have generally lower mean MRSS than that of non-completers for nearly the entire duration of the trial. This suggests that the probability for patients dropping out is likely to be related to their high MRSS (signalling more severe

disease), and thus these values are not MCAR. However, graphical techniques to further distinguish between MAR and NMAR are not available.

There was a difference in the proportions in the completers and non-completers groups who received bovine collagen treatment. Among completers, 43.4% were in the collagen group and among non-completers, 61.8% were in the collagen group ($p=0.037$). This suggests that the probability for patients dropping out is likely to be related to their treatment status. If the missing data mechanism is MCAR, characteristics between completers and non-completers should be similar. Since both MRSS and treatment status are different between completers and non-completers, missing data mechanism is not MCAR. On the other hand, it is not possible to verify whether this data follows MAR or NMAR based on data, since we have no way of examining unobserved data.

Table II shows the estimated difference and p -values for testing equality of changes in MRSS in late phase patients between the two groups at 12 and 15 months using 11 different methods of analysis. At 12 months, complete-case analysis shows a significant difference with the estimated decrease in MRSS at 12 months 3.63 points lower in the bovine CI group than in the placebo group (p -value=0.040). However, none of the other approaches indicate significant differences between the two groups at the 5% significance level.

At 15 months, the last observation carried forward (LOCF) approach showed no difference (p -value=0.066) while all the other methods demonstrated a significant difference between CI and placebo. This indicates that depending on the imputation method for missing data, one might draw differing conclusions about the efficacy of bovine CI treatment. Furthermore, when examining the methods which did show a difference, the p -values varied substantially, and thus the confidence with which one drew conclusions also varied depending on the missing data method used.

The estimates of the difference in decrease of the MRSS score at both 12 and 15 months between the two groups are much smaller with the LOCF method. However, we view these results with

Table I. Methods to analyse longitudinal clinical trial data with missing data.

Category	Dropping or Ignoring Missing Cases	Single Imputation	Multiple Imputation	Likelihood-based method	Joint likelihood-based models for data and missingness	Pseudo-likelihood-based methods or GEE
Methods	(i) Complete case analysis-analyses only patients with no missing values (ii) Available case analysis-analyses data as they are without filling in missing values	(i) Last Observation Carried Forward (LOCF)-this means filling in each missing value at each analysis by the last available data from the patient; (ii) overall mean imputation - this means filling in all missing values at end point by a single value averaged over the whole cohort; (iii) conditional mean imputation- this means filling in each missing value using a value averaged over specific characteristics of the patient, for example, using the predicted value in a regression model	Joint modelling imputation - assumes a model and fills in each missing value repeatedly using methods such as those described in the previous column so that one can account for the variability of the filled-in value (ii) conditional modelling imputation -similar to above but filled-in values are now random draws from a constructed distribution that already incorporated information from all observed data	(i) Mixed effects models-allows between and within patient variability to be accounted for in the model for predicting what the missing value might have been; (ii) repeated measures models-this means use a regression model and predicting missing values using correlations of the serial responses within each patient	(i) Combination of two models: a mixed effects model for data and a logistic regression for missingness- two sets of models are required: one that describes the supposed missing mechanism and the other for filling in the missing value using methods such as those just described in the previous column.	(i) Generalised Estimating Equations (GEE)-a general and very robust technique for analysing a longitudinal study after missing values are filled in using multiple imputation; assumptions required in the statistical model to do the analysis are minimal
Positives	Extremely simple	Relatively simple to perform	Accounts for uncertainty of entered missing data	Accounts for within-subject correlation. Can also be used with binary and count data. Use all available information.	Explicit handling of NMAR data	Does not require that data is normally distributed. Estimates are consistent and accurate when there are a large number of patients. Results remain valid even when we mis-specify the correlations of responses within patients
Negatives	Requires data to be missing completely at random (see text for definition) plus requires large amounts of data	Ignores uncertainty in imputed values	Requires that missing data be Missing At Random (see text for definition)	Requires that missing data be Missing At Random (see text for definition).	Cannot tell if the Not Missing At Random (see text for definition) assumptions are correct using the available data.	Requires that missing data be Missing Completely At Random (see text for definition)
Example		When HAQ-DI of patient at any visit is missing, overall mean imputation replaces missing values with mean of HAQ-DI at that specific visit across all patients	Missingness of MRSS is spread across visits and patients. Select a set of covariates where data is filled in(eg joint counts, HAQ-DI, FVC, disease duration) upon which to base the missing data. An algorithm creates a specified number of new datasets populated with imputed values for both MRSS and covariates. The new datasets are analysed separately using complete case methods, and results are then combined to yield a single set of results.	Missingness of MRSS is spread across visits and patients. For example, one selects a set of covariates to regress MRSS upon. Specify random intercept (value when MRSS is 0) to account for the patient's specific covariate when MRSS is 0. May specify additional random-effects to allow for the differences among patients.	Use mixed effects model for data. Assume a regression model for a binary outcome variable where probability of missingness can depend on covariates, observed data points and unobserved data points. Specifically, we assume that whether the observation for the patient at a particular time is missing or not depends on a specific set of covariates and the observed and unobserved responses at an earlier time via a random effects logistic model. For example, the missing MRSS is assumed to be dependent on disease duration, distance from clinic, drugs used in the past, FVC, etc at baseline and then used with a random effects logistic model to predict the missing MRSS at the missing data-point.	Missingness of MRSS is spread across visits and patients. One then selects a set of covariates (for example, duration of disease, FVC, drugs etc) to regress MRSS upon and specifies a structure for the within-subject correlations for all patients, (for example, unstructured, exchangeable). Since a Missing Completely At Random assumption is required, GEE can be directly applied for missing data when this assumption is satisfied. If missing data mechanism is not missing at random, multiple imputation is performed to impute missing values, and the GEE analysis is performed for imputed data.

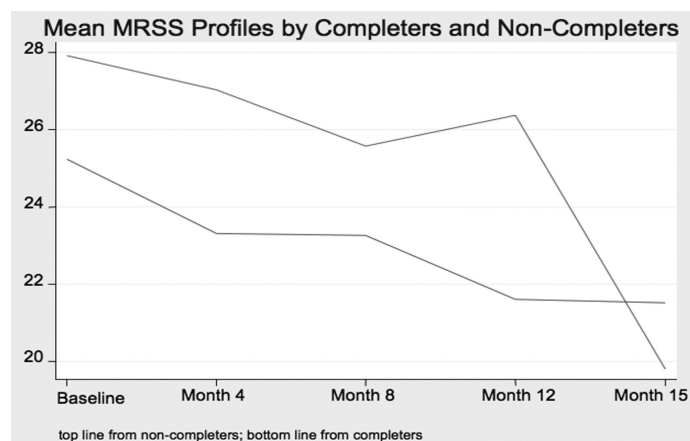


Fig. 1. Mean MRSS profiles for completers and non-completers from baseline to Month 15.

Table II. Estimates and *p*-values for testing changes in MRSS in late phase patients between the two groups at 12 (15) months using eleven different methods of analysis.

Number	Statistical method	12-month Difference* (<i>p</i> -value)	15-month Difference (<i>p</i> -value)
1	Complete case	-3.63 (0.040)	-4.86 (0.005)
2	Available case	-2.67 (0.103)	-4.97 (0.005)
3	LOCF	-1.71 (0.249)	-2.76 (0.066)
4	Mixed effects (with random intercept time-specific variances)	-2.65 (0.066)	-4.35 (0.004)
5	Repeated measures (with unstructured covariance matrix)	-2.74 (0.090)	-4.11 (0.013)
6	NMAR (Carpenter's model)	-2.68 (0.075)	-4.32 (0.003)
7	NMAR (Boscardin's model)	-2.21 (0.106)	-4.19 (0.005)
8	Multiple Imputation (Complete case)	-2.49 (0.150)	-4.25 (0.020)
9	Multiple Imputation (Mixed effects)	-2.49 (0.106)	-4.25 (0.010)
10	Multiple Imputation (Repeated measures)	-2.49 (0.144)	-4.25 (0.016)
11	Multiple Imputation (GEE)	-2.49 (0.138)	-4.25 (0.018)

*Difference indicates mean change in MRSS at 12 or 15 months from baseline for the bovine group; mean change in MRSS at 12 or 15 months from the baseline for the control group.

some skepticism because we already observed from Figure 1 a decreasing trend in MRSS over time for both groups and we know that the LOCF method can give biased estimates when data show decreasing or increasing trend over time (Tang *et al.*, 2005). Results from the analysis based on the LOCF method would therefore seem inappropriate for this data set.

Complete case analysis, LOCF and available case analysis require data to be MCAR and for LOCF, it also assumes that all missing data after the last observed data have the same values as the last observed value. However, as shown earlier, the missing data mechanism for this data is not MCAR, since completers and non-completers have different MRSS scores and differences in the proportion of patients receiving the bovine CI treatment. Therefore, these two methods would also not be most appropriate methods to handle missing values in this data.

The two MAR-valid approaches (mixed effects and repeated measures model) provide similar estimates and *p*-values for the difference between the two groups. The four approaches based on multiple imputation and the two NMAR techniques did not give substantially different conclusions, suggesting that the results are robust to the MAR assumption. We note that the mean estimates from the four multiple imputation methods are the same because the data become balanced after imputation and we fitted the saturated model. However, their estimates for the covariances of the estimated model parameters are different because the analysis models are different.

Discussion

In this article, we discuss various statistical methods of handling missing data in a two-arm randomised clinical trial and apply them to analyse a multi-centre randomised placebo controlled two-arm SSc trial as an example.

The analysis based on completers alone (complete case analysis and available case analysis) and the last observation carried forward (LOCF) methods require underlying assumptions which are rarely met in practice. For example, both methods require large amounts of data and, for LOCF, one must also assume that the imputed data are correct. Thus, these two methods will often lead to inherent biases. In our example, they gave different conclusions, especially at 12 months.

Multiple imputation, mixed effects, and repeated measures models are more sophisticated and realistic attempts to handle missing data in a clinical trial. They try to account for the differences among patients and account for patient's specific response patterns. This is likely to result in analyses that are less biased. However, the underlying assumption that the missing data is only directly related to the observed characteristics may well not be true. For example, the patient could not come in for a visit because there was a transportation problem or had a family emergency, but such information is rarely included in the observed data, thereby making it harder to ascertain the true nature of the missingness.

The joint likelihood based model combines the mixed effect model and logistic regression model to explicitly handle data not missing at random and so it is more realistic and potentially takes an additional step toward decreasing bias. Its weakness is that there is no way to test whether the data are truly not missing at random. Frequently for the case of a randomised controlled trial, informal methods are sometimes employed to ascertain whether data are not missing at random. For example in our data set, if one is willing to assume that HAQ-DI and MRSS are correlated, it is safe to assume that MRSS data is NMAR with respect to the HAQ-DI. From Table II, one observes that using the model for analysing NMAR data, we obtained results that are consistent with those from the repeated measures, mixed effects models, and multiple imputations. The confidence of the results for these methods also appeared greater than for the other models.

One can see that the various methods for

analysis all have positive and negative aspects (Table I), making it important that a specific primary statistical analysis be chosen well in advance of data completion date. Otherwise, a form of “cherry picking” could occur, and this can significantly compromise the credibility of the results.

Consistent results from these approaches towards missing data can provide assurance that the missing data do not have a major effect on the primary conclusion. They should be used, appropriately, in predefined, sensitivity analyses. On the other hand, investigators need to resist any temptation to choose, *ex post facto*, the particular analysis that produces the most significant results.

A major difficulty of the analysis of missing data is that the missing data mechanism is often not known and there is a limitation to validate it in real data. Therefore, this article recommends that the data analyst should always perform several different analyses and compare results. If results are seriously different, it is better to choose the one based on weaker assumptions. In the bovine CI trial, eleven different approaches provide quite different results. However, the difference mainly comes from results based on two groups of analysis methods, where within the group results are similar. One group consists of complete case analysis, available case analysis, and last observation carried forward method and the other group consists of the eight other methods assuming either MAR or NMAR. Since the former group assumes MCAR and Figure 1 indicates that MCAR is not an appropriate assumption, the three methods in the first group should be avoided. On the other hand, the eight approaches that assumed data are either MAR or NMAR provided consistent results. It follows that if any one or more of the methods in the second group were predefined plans to analyse the data, they would have been appropriate.

A motivation for this paper comes from our observation that many clinical trials for rheumatic diseases are analysed by less experienced statisticians, sometimes supervised by a PhD-level statistician. Some less experienced statisticians are not trained in handling missing data using the various ap-

proaches described herein, including understanding the pros and cons of each method, and statistical tools for verifying the assumed missing data mechanism. We hope that this paper informs rheumatologists of missing data issues and enables them to ask appropriate questions before arriving at the analysis conclusion. Doing so will avoid “cherry picking” the results from one analysis just because of its conclusion. In this article, we recommend trying sensitivity analysis of the conclusion based on various methods. If all approaches provide similar results, it may indicate that the missing data mechanism is MCAR. On the other hand, if all results provide different results as in our example, we recommend that the conclusion be based on an analysis that relies on the NMAR assumption.

Our analysis focused on one variable, MRSS. When data include many variables, missingness on different variables may arise from different missing data mechanisms. In this case, it is acceptable to choose different approaches for analysing missing data from different variables. For example, if the missing data mechanism for the first variable is MCAR and the missing data mechanism for the second variable is NMAR, then it is appropriate to analyse or impute data using MCAR assumption while analysing the second variable using the NMAR assumption. On the other hand, it is often hard to verify the missing data mechanism for each variable. One option is to choose one approach to handle all variables. For our trial, we analysed both variables under the NMAR assumption. This approach is preferred since it provides a consistent approach to all variables and an appropriate analysis under NMAR would not be incorrect even under MCAR.

The clinician investigator and the statistician need to carefully discuss and agree upon all of the underlying factors and assumptions for an imputation model which accounts for missing data, before analysis begins. Ideally, in fact, the analyses, including the approaches to anticipated missing data, should be included in the initial protocol, before the actual study commences.

In conclusion, using the Bovine CI Study in SSc, we demonstrate the ef-

fects of using different techniques for handling missing data, including novel approaches to account for MAR and NMAR data. The last observation carried forward approach, although widely used, may give biased results and is often inappropriate. Missing data can have major impact on the statistical inference and so the analytic technique needs to be carefully considered and chosen *a priori*, although sensitivity analyses can reveal problems or support the primary analysis.

Our recommendations for future RCTs include:

1. Before undertaking the trial, decide which method for imputing missing data would be most appropriate. Usually it is one which works under the MAR and NMAR assumptions.
2. After the trial, check your approach by ascertaining whether data is MCAR. This can be assessed graphically in patients who completed the trial vs. those who did not complete the trial, using the primary outcome and, if desired other key secondary outcome measures. If the data are not MCAR, one can use the predefined approaches as valid sensitivity analyses and report the final results.

References

1. BOSCARDIN JW, YAN X, WONG WK: A reanalysis of a longitudinal scleroderma clinical trial using informative dropout models. *J Stat Plan Infer* 2007; 137: 3848-58.
2. CARPENTER J, POCOCK S, LAMM CJ: Coping with missing data in clinical trials: a model-based approach applied to asthma trials. *Stat Med* 2002; 21: 1043-66.
3. CARPENTER J, KENWARD M, EVANS S, WHITE I: Last observation carry-forward and last observation analysis. *Stat Med* 2004; 23: 3241-4.
4. LITTLE RJA: A test of missing completely at random for multivariate data with missing values. *Am Stat* 1988; 83: 1198-202.
5. MOLENBERGHS G, KENWARD M: *Missing Data in Clinical Studies*, Wiley: New York. 2007.
6. POSTLETHWAITE AE, WONG WK, CLEMENTS P *et al.*: A multicenter, randomized, double-blind, placebo-controlled trial of oral type I collagen treatment in patients with diffuse cutaneous systemic sclerosis: i. oral type I collagen does not improve skin in all patients, but may improve skin in late-phase disease. *Arthritis Rheum* 2008; 58: 1810-22.
7. RUBIN DB: Inference and missing data. *Biometrika* 1976; 63: 581-92.
8. TANG L, SONG J, BELIN TR, UNÜTZER J: A comparison of imputation methods in a longitudinal randomized clinical trial. *Stat Med* 2005; 24: 2111-28.