

SLE clinical trials: impact of missing data on estimating treatment effects

Mimi Kim ¹, Joan T Merrill,² Cuiling Wang,¹ Shankar Viswanathan,¹ Ken Kalunian,³ Leslie Hanrahan,⁴ Peter Izmirly⁵

To cite: Kim M, Merrill JT, Wang C, *et al.* SLE clinical trials: impact of missing data on estimating treatment effects. *Lupus Science & Medicine* 2019;**6**:e000348. doi:10.1136/lupus-2019-000348

Received 8 July 2019
Revised 23 August 2019
Accepted 28 August 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York, USA

²Arthritis & Clinical Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma, USA

³Rheumatology, UCSD, San Diego, California, USA

⁴Research and Education, Lupus Foundation of America, Washington, District of Columbia, USA

⁵Medicine, Division of Rheumatology, New York University School of Medicine, New York City, New York, USA

Correspondence to

Dr Mimi Kim; mimi.kim@einstein.yu.edu

ABSTRACT

Objective A common problem in clinical trials is missing data due to participant dropout and loss to follow-up, an issue which continues to receive considerable attention in the clinical research community. Our objective was to examine and compare current and alternative methods for handling missing data in SLE trials with a particular focus on multiple imputation, a flexible technique that has been applied in different disease settings but not to address missing data in the primary outcome of an SLE trial.

Methods Data on 279 patients with SLE randomised to standard of care (SoC) and also receiving mycophenolate mofetil (MMF), azathioprine or methotrexate were obtained from the Lupus Foundation of America-Collective Data Analysis Initiative Database. Complete case analysis (CC), last observation carried forward (LOCF), non-responder imputation (NRI) and multiple imputation (MI) were applied to handle missing data in an analysis to assess differences in SLE Responder Index-5 (SRI-5) response rates at 52 weeks between patients on SoC treated with MMF versus other immunosuppressants (non-MMF).

Results The rates of missing data were 32% in the MMF and 23% in the non-MMF groups. As expected, the NRI missing data approach yielded the lowest estimated response rates. The smallest and least significant estimates of differences between groups were observed with LOCF, and precision was lowest with the CC method. Estimated between-group differences were magnified with the MI approach, and imputing SRI-5 directly versus deriving SRI-5 after separately imputing its individual components yielded similar results.

Conclusion The potential advantages of applying MI to address missing data in an SLE trial include reduced bias when estimating treatment effects, and measures of precision that properly reflect uncertainty in the imputations. However, results can vary depending on the imputation model used, and the underlying assumptions should be plausible. Sensitivity analysis should be conducted to demonstrate robustness of results, especially when missing data proportions are high.

INTRODUCTION

A frequent problem in clinical trials of new treatments for SLE and other diseases is missing outcome data due to participant dropout, loss to follow-up, skipped visits and other factors. This can result in diminished statistical power to differentiate effective experimental therapies from standard of care

(SoC), as well as biased estimates of treatment effects. In 2010, the Food and Drug Administration (FDA) commissioned a report by the National Research Council of the National Science Foundation on the proper handling of missing data.¹ However, these recommendations are not being widely incorporated into clinical practice. A review of randomised clinical trials published in 2013 in leading medical journals found that 95% reported some missing outcome data, yet the issue was not properly addressed in over 70% of the cases. A 2018 editorial in the *Annals of Internal Medicine* also highlighted the large gap between the growing availability of new techniques for handling missing data and the application of these modern methods to actual studies, and emphasised the importance of wider dissemination of the information to the research community.² The aim of this paper is to examine and compare the strengths and limitations of current and alternative methods for addressing missing data in SLE trials with a particular focus on multiple imputation (MI).

When choosing a missing data method, the aim should be to maximise use of the available information in the trial, minimise bias in results, and obtain estimates of the precision of the results that properly reflect the uncertainty in any values that are imputed for the missing data. One also needs to consider whether the method's assumptions regarding the mechanisms that caused the missing information are reasonable. These mechanisms are typically classified into three categories: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). To differentiate them in the context of a lupus trial, we assume the outcome of interest is SLE Responder Index-5 (SRI-5) at 52 weeks, a composite endpoint used in several recent trials.^{3,4} When SRI-5 data are missing but the probability that these are missing does not depend on any observed or unobserved factors, then SRI-5 is considered

to be MCAR. In contrast, suppose that SRI-5 missing data rates are higher in patients with SLE Disease Activity Index (SLEDAI) ≥ 10 at baseline. The MCAR assumption would clearly not be satisfied here since baseline disease severity predicts the probability that SRI-5 is missing at 52 weeks. But if this probability depends just on the baseline SLEDAI score and not on the unobserved value of SRI-5 or other factors, then the missingness would be considered MAR after conditioning or accounting for SLEDAI score when analysing treatment effects. In general, MAR holds when the probability that data are missing depends only on observed factors. Finally, SRI-5 data are MNAR if the missingness depends on the missing value directly, that is, SRI-5 is more likely to be missing in those who would have been non-responders at week 52 in our case or because of unmeasured factors.

The most common approaches for dealing with missing data in clinical studies are (1) complete case analysis (CC), which we define here as excluding from the analysis subjects with missing values for the outcome of interest; (2) last observation carried forward (LOCF), which fills in for the missing data the subject's last observed value for the outcome; and (3) non-responder imputation (NRI), which assumes all missing outcomes are failures or non-responses. In the CC approach, results will be unbiased either if those without missing data values are a random subset of the original study population (MCAR), or if the missing data are MAR and the analysis has properly controlled for all the variables which affect the probability of missing data. For example, if the probability that SRI-5 is missing depends just on baseline SLEDAI score, then to obtain unbiased results with the CC method, treatment effects should be estimated with a statistical approach that can control for SLEDAI, for example, logistic regression. However, whether the missing data are MCAR or MAR, the available sample size and power of the study will be reduced in the CC approach because of the missing data.

Many SLE trials use LOCF or NRI or both in the same trial to deal with missing data.^{3–5} These are known as 'single imputation' methods because the missing data are imputed or filled in once with either the last observed value or a value corresponding to non-response, respectively, to generate a complete data set that includes all randomised subjects for analysis. However, LOCF can bias results in the positive or negative direction if the outcome is changing over time or patients withdraw because of deteriorating health.⁶

With the NRI approach, the estimated response rates in each treatment group will clearly be attenuated relative to the true rates, with the degree of attenuation exactly equal to the missing data rate. For example, if the true response rate in a treatment arm is 40% and the proportion with missing data is 20%, then the NRI-based response rate will be biased downwards by 20%, that is, 32% instead of 40% in that arm. But as with LOCF and CC, the estimated treatment effect or difference in response rates between treatment arms using NRI can be biased in either direction.

In addition to potentially yielding biased results, single imputation methods like LOCF and NRI assume that the 'guesses' used to fill in the missing outcomes are the true values. The resulting SE and width of the corresponding CI for the treatment effect will be smaller than they should be.⁷ The 2010 FDA-commissioned guidelines on missing data strongly discouraged the use of these single imputation methods, and instead recommended using alternative methods that better account for the uncertainty in values used to impute missing data.¹

MI is a well-established and flexible model-based technique for filling in the missing values that properly takes into account the uncertainty in the imputation process.^{8,9} MI has been applied to studies of rheumatoid arthritis⁹ and osteoarthritis,¹⁰ but not, to our knowledge, to address missing outcome data in an SLE trial. In this paper, we explore the use of MI to handle missing SRI data from a clinical trial where the outcome measured at the final visit is of primary interest, and compare it with the CC, LOCF and NRI approaches. The methods are illustrated in an example to assess whether among patients assigned to the SoC arm in a 52-week SLE trial response rates at the end of follow-up differ by type of patients' background immunosuppressant treatment during the trial.

METHODS

Multiple imputation

The MI procedure involves the following three steps: imputation, analysis and pooling.

Imputation step

Missing values are imputed using predictions for the missing outcome that are generated from a statistical model. Note that this imputation model is distinct from the one that is used to evaluate treatment effects in the main analysis. Two common approaches for generating the imputations are Markov chain Monte Carlo algorithm, which assumes that the variables in the imputation model jointly follow a multivariate normal distribution, and MI based on chained equations (also known as fully conditional specification). We chose chained equations to impute the missing data since the method can handle both missing continuous and categorical variables as well as arbitrary missing data patterns.¹¹ In this method, a separate regression model is specified for each variable that has missing data, for example, linear regression for continuous variables and logistic regression for binary variables such as SRI-5. After using single imputation to initially fill in any missing data, the regression models are sequentially fit to each variable using only the observed data for the target (dependent) variable and the observed and imputed values for the other variables that are used as predictors in the model. The fitted regression is used to generate new predictions or imputations to fill in missing data in the target variable. After cycling through all the variables in this manner, the procedure is repeated a number of times to stabilise the results before generating

one complete data set. Further technical details of the chained equations method are described elsewhere.^{11–14}

Since the MI method requires the MAR assumption, the imputation model for the primary outcome in an SLE trial should include as predictors all the variables that will be adjusted for in the main analysis to evaluate treatment effects (eg, any stratification factors such as SLEDAI score or anti-double stranded DNA (anti-dsDNA)), and the variables that might predict both the value of the missing outcome and the probability that the outcome is missing,^{12–15} such as disease activity measures obtained at earlier visits. Including a few variables in the imputation model that are highly correlated with the outcome is better than having many with low correlations because these variables are often intercorrelated, and the incremental benefit of adding them to the model is small.¹⁶ However, while including variables that are not related to the variable being imputed in the imputation models may slightly decrease efficiency, it should not cause bias.^{17–18}

To take into account the uncertainty in the imputed values, the imputation step is performed multiple times (M) so that multiple plausible values for the missing data are generated, resulting in M complete data sets. A common rule of thumb is to set M equal to at least the percentage of incomplete cases, for example, if 30% have missing data, then generate $M=30$ complete data sets.¹²

Analysis step

Each of the M complete data sets is analysed separately using the main statistical method (eg, logistic regression) to obtain M different estimates of the treatment effects, such as ORs or differences in response rates, and corresponding SEs.

Pooling step

The M sets of results are then combined using Rubin's rules¹⁹ so that the final MI estimate of the treatment effect is a simple average of the M treatment effect estimates. The corresponding variance of the estimated treatment effect takes into account the within-imputation and between-imputation variances.

CDAI data

To illustrate the missing data methods with an SLE trial, we used data from the Collective Data Analysis Initiative Database of the Lupus Foundation of America (LFA-CDAI), which was established so that data from the placebo/SoC arms of previous SLE clinical trials can be used to improve the design and conduct of future studies. Since we did not have access to data from the experimental treatment arm, our goal was to apply the different missing data methods in an unadjusted comparison of the SRI-5 response rates at 52 weeks between SoC patients on mycophenolate mofetil (MMF group) versus other immunosuppressants (non-MMF group) at baseline. Data from 279 patients with SLE in the LFA-CDAI database without nephritis who were receiving MMF, azathioprine or methotrexate at entry into the trial were included in

the analysis. Reasons for missing data were not available in the data set.

Imputing SRI-5

This variable is a composite outcome that is defined by a decrease in SLEDAI score by at least five points since baseline, no new British Isles Lupus Assessment Group (BILAG) index score of A (severe) organ activity, no more than one new BILAG B (moderate) organ domain score, and no worsening in Physician Global Assessment (PGA) score (increase <0.3 points on a 3-point scale) from baseline. It was unclear whether missing SRI-5 should be imputed directly at the composite level or if the individual components should be imputed first and then SRI-5 derived based on those imputed values. Therefore, we performed MI using three different MI models for SRI-5 at 52 weeks. The first imputation model (*MI-1*) was a logistic regression model with SRI-5 specified as the binary outcome. Since none of the baseline characteristics was observed to be significantly different between patients with and without missing SRI-5 data at 52 weeks (table 2), the predictor variables were selected based on clinical considerations and prior studies that found they were associated with disease status and SRI-5 response duration during follow-up.²⁰ These variables included MMF use, SRI-5 at four earlier visits (weeks 12, 24, 36 and 44) which were moderately to highly correlated with SRI-5 at 52 weeks (correlation coefficient=0.5–0.8), and the following variables: race, baseline values of SLEDAI, PGA, BILAG score, protein to creatinine ratio and anti-dsDNA. The second model (*MI-2*) included the same demographic and clinical variables as the first one, but rather than directly impute SRI-5 we first filled in the missing values for the individual component measures at week 52 and the four earlier visits using the chained equations approach, and then determined SRI-5 status based on those imputed values. For simplicity, rather than impute all of the organ-specific BILAG ordinal scores, the binary variable for the BILAG component of SRI-5 (no new BILAG A and fewer than 2 BILAG Bs compared with baseline) was directly imputed instead. The third imputation model for SRI-5 at 52 weeks (*MI-3*) included the same variables as in MI-1 as well as SRI-5 at eight additional visits (weeks 4, 8, 16, 20, 28, 32, 40 and 48) to explore the impact of including a large number of predictors that are highly correlated with each other and with the outcome in the same imputation model. For each imputation approach, 40 imputed data sets were generated and analysed to estimate the difference in response rates between the MMF and non-MMF groups. MI results were compared with those from using the CC, LOCF and NRI methods.

Finally, MI depends on the MAR assumption, but MAR cannot be distinguished from MNAR with observed data. Sensitivity analyses are therefore strongly recommended to assess the robustness of findings to departures from the MAR assumption. In our sensitivity analysis, we used a pattern-mixture modelling framework to perform a 'stress

Table 1 Missing data patterns in CDAI example data

| Pattern | 12 weeks | 24 weeks | 36 weeks | 44 weeks | 52 weeks | Frequency (%) |
|---------|----------|----------|----------|----------|----------|---------------|
| 1 | X | X | X | X | X | 204 (73) |
| 2 | X | X | X | X | O | 2 (0.72) |
| 3 | X | X | X | O | X | 3 (1.08) |
| 4 | X | X | X | O | O | 4 (1.43) |
| 5 | X | X | O | O | O | 8 (2.87) |
| 6 | X | O | X | X | X | 1 (0.36) |
| 7 | X | O | X | O | X | 1 (0.36) |
| 8 | X | O | O | O | O | 31 (11.11) |
| 9 | O | X | X | X | X | 1 (0.36) |
| 10 | O | O | O | O | O | 24 (8.60) |

CDAI, Collective Data Analysis Initiative Database; O, missing; X, observed.

test' of the MAR assumption. We systematically adjusted the imputation model used to fill in the missing data in the non-MMF group so that the probability of SRI-5 response became progressively lower than that of subjects who had non-missing data, resulting in a MNAR pattern. The point at which the conclusion about the treatment groups is reversed is known as the tipping point.^{21 22}

All analyses were conducted in SAS V.9.4. The MI approach was performed using the PROC MI procedure to generate the MIs and PROC MIANALYZE to pool the results across the complete data sets.

RESULTS

Sixty patients were in the MMF-treated group and 219 patients were treated with different medications. [Table 1](#) shows the missing data pattern in longitudinal measures of SRI-5 at weeks 12, 24, 36, 44 and 52. In this subset of visits, the majority of patients showed a monotonic missing data pattern in which once data are missing at a particular visit, data at all subsequent visits are also missing. Only six patients (2%) had a missing value followed by an observed value at a future visit, indicating that intermittently missed visits were not common. Missing data rates for SRI-5 at 52 weeks were 32% and 23% in the MMF and non-MMF groups, respectively. SLEDAI, PGA and BILAG each had overall missing data rates of 24%; only 8% had partial information for these indices, so that the majority of patients had data on all or none of the individual components of SRI-5 at 52 weeks.

The baseline characteristics of patients with missing and non-missing SRI-5 data at 52 weeks are summarised and compared in [table 2](#). Disease activity measures, disease duration, age, race, steroid use and laboratory values were not significantly different between those with and without missing SRI-5 data at 52 weeks.

The CC, LOCF, NRI and MI approaches were each applied to handle the missing data and obtain estimates of the 52-week SRI-5 response rates in the MMF and non-MMF groups, between-group difference in response rates, corresponding 95% CIs and p values ([table 3](#)). With

the CC approach, the SRI-5 response rates were 46.8% in non-MMF and 29.3% in MMF, corresponding to a difference of 17.5% ($p=0.043$). As expected, CIs were widest when analyses were based only on the available data, reflecting the loss of precision that results when the sample size is reduced because of missing data. LOCF produced the smallest and least significant estimates of between-group differences in SRI-5 response rates at 52 weeks (10.6%; $p=0.13$). LOCF assumes that the missing outcome is equal to the last observed value of the outcome regardless of when it occurred. This is an unrealistic assumption in our case since in a separate longitudinal analysis that we conducted, SRI-5 rates were observed to increase over time in both the MMF and non-MMF groups, but at a faster rate in the non-MMF group. The NRI method yielded the lowest estimated response rates in each treatment group since all missing outcomes are assumed to be non-responses. The NRI-estimated between-group difference (16.1%; $p=0.019$) was slightly smaller than the CC estimate.

With multiple imputation models MI-1 (directly imputing SRI-5) and MI-2 (deriving SRI-5 from imputed values of individual components), differences in response rates between the MMF and non-MMF groups were magnified and more statistically significant compared with the other missing data approaches (MI-1: 19.1%, $p=0.010$; MI-2: 19.0%, $p=0.011$). However, CIs with the two MI models were wider than with the LOCF and NRI methods since the MI procedure appropriately takes into account the uncertainty in the imputed values. Similarity in results between MI-1 and MI-2 suggests that imputing at the composite rather than individual component level should be sufficient unless the individual components are themselves of interest and will be separately analysed anyway. Estimated between-group differences from MI-3, which was identical to MI-1 but included SRI-5 status at eight more visits, were smaller and less significant (17.4%; $p=0.026$) than the results from the other MI models. This may reflect the lack of precision that can result when an unnecessarily large number of highly correlated variables are simultaneously included in the imputation model.

Table 2 Baseline characteristics of subjects with and without missing SRI-5 data at 52 weeks

| Variable | Missing SRI-5 at 52 weeks | | P value |
|--------------------------------------|---------------------------|---------------|---------|
| | Yes (n=69) | No (n=210) | |
| MMF, n (%) | 19 (27.5) | 41 (19.5) | 0.16 |
| White race, n (%) | 42 (60.9) | 119 (56.7) | 0.54 |
| Age (years) | 36.9 (12.31) | 39.0 (11.82) | 0.21 |
| Disease duration (years) | 4.68 (5.81) | 4.90 (7.71) | 0.60 |
| Prednisone dose (mg/day) | 10.0 (12.5) | 10.0 (10.0) | 0.15 |
| BILAG score | 15.75 (8.38) | 14.75 (7.11) | 0.33 |
| SLEDAI | 10.68 (4.31) | 10.60 (3.70) | 0.89 |
| PGA | 46.50 (19.59) | 47.09 (16.25) | 0.82 |
| White cell count ($\times 10^9/L$) | 5.68 (2.51) | 5.90 (2.49) | 0.53 |
| Lymphocytes ($\times 10^9/L$) | 1.14 (0.52) | 1.24 (0.61) | 0.23 |
| Neutrophils ($\times 10^9/L$) | 3.71 (2.61) | 3.64 (2.46) | 0.92 |
| CD19 (%) | 11.45 (13.5) | 9.10 (11.1) | 0.14 |
| Haemoglobin (g/L) | 122.0 (13.4) | 123.7 (15.1) | 0.41 |
| C3 (g/L) | 0.97 (0.36) | 1.00 (0.31) | 0.61 |
| C4 (g/L) | 17.2 (15.7) | 15.1 (12.3) | 0.39 |
| Anti-dsDNA (IU/mL) | 120.5 (245.5) | 69.0 (288.0) | 0.99 |
| IgG (g/L) | 14.99 (5.21) | 15.70 (5.49) | 0.35 |
| IgM (g/L) | 0.91 (0.74) | 0.99 (0.78) | 0.27 |
| IgA (g/L) | 2.63 (1.69) | 2.97 (1.56) | 0.12 |
| Anti-SM (U/mL) | 5.5 (21.5) | 5.5 (2.8) | 0.51 |
| Anti-RNP (U/mL) | 5.5 (82.0) | 5.5 (88.0) | 0.25 |
| Anti-SSA/Ro (U/mL) | 5.5 (128.5) | 5.5 (102.0) | 0.77 |
| Anti-SSB/La (U/mL) | 13.75 (29.03) | 21.96 (38.52) | 0.06 |
| Protein to creatinine ratio | 0.07 (0.12) | 0.06 (0.11) | 0.49 |

Values are mean (SD) or median (IQR) depending on distribution of the data.

P values were based on t-tests or Wilcoxon tests for continuous variables, and χ^2 tests for categorical variables.

anti-dsDNA, anti-double stranded DNA; anti-RNP, anti-ribonucleoprotein; anti-SM, anti-Smith ; anti-SSA/Ro, anti-Sjögren's syndrome related antigen A; anti-SSB/La, anti-Sjögren's syndrome type B; BILAG, British Isles Lupus Assessment Group; MMF, mycophenolate mofetil; PGA, Physician Global Assessment; SLEDAI, SLE Disease Activity Index; SRI-5, SLE Responder Index-5.

With the exception of LOCF, all the approaches led to the same conclusion that SRI-5 response rates were significantly lower in patients on MMF. Since the MI procedure depends on the MAR assumption and there

Table 3 SRI-5 response rates at 52 weeks on SoC by immunosuppressant use

| Missing data approach | Non-MMF (%) | MMF (%) | Difference (%) | 95% CI | P value |
|-----------------------|-------------|---------|----------------|--------------|---------|
| CC | 46.8 | 29.3 | 17.5 | 1.7 to 33.3 | 0.043 |
| LOCF | 40.6 | 30.0 | 10.6 | -2.7 to 23.9 | 0.13 |
| NRI | 36.1 | 20.0 | 16.1 | 4.1 to 28.0 | 0.019 |
| MI-1* | 47.6 | 28.5 | 19.1 | 4.6 to 33.5 | 0.010 |
| MI-2* | 46.0 | 27.0 | 19.0 | 4.3 to 33.6 | 0.011 |
| MI-3* | 48.2 | 30.8 | 17.4 | 2.0 to 32.7 | 0.026 |

*n=40 imputed data sets. MI-1: imputation model includes MMF status, race, baseline values of SLEDAI, PGA and BILAG score, protein to creatinine ratio, anti-dsDNA, and SRI-5 at 12, 24, 36, 44 and 52 weeks; MI-2: imputation model same as MI-1 but separately imputing components of SRI-5 (SLEDAI, BILAG and PGA); MI-3: same as MI-1 but additionally includes SRI-5 at all 13 visits. anti-dsDNA, anti-double stranded DNA; BILAG, British Isles Lupus Assessment Group; CC, complete case analysis; LOCF, last observation carried forward; MI, multiple imputation; MMF, mycophenolate mofetil; NRI, non-responder imputation; PGA, Physician Global Assessment; SLEDAI, SLE Disease Activity Index; SoC, standard of care; SRI-5, SLE Responder Index-5.

is no way to know with the observed data whether MAR truly holds, we performed additional sensitivity analysis to assess how much the data can deviate from the MAR assumption before the main conclusion that SRI-5 rate at 52 weeks is significantly lower in the MMF-treated patients is reversed. This tipping point did not occur until the odds of having an SRI-5 response among those with missing outcomes in the non-MMF group were adjusted in the imputation model to be fourfold lower than in subjects with non-missing outcomes, providing additional support for the stability of the overall conclusion.

DISCUSSION

Potential advantages of applying the MI approach over the CC, LOCF or NRI in SLE trials include reduced bias in treatment effect estimates and measures of precision that properly reflect uncertainty in the imputations. In our example, MI resulted in larger and more significant between-group differences in response rates compared with CC, LOCF and NRI; this trend would not necessarily hold in all data sets.

MI is particularly advantageous when strong predictors of the missing outcome are available. When the primary endpoint in the SLE trial is SRI-5 or other disease outcome at the last follow-up visit, imputation models for the missing outcome should at the very least include treatment arm, randomisation stratification factors that will be adjusted for in the main analysis, and outcomes measured at earlier visits if they are highly correlated with subsequent outcomes. However, including data from all visits in the same imputation model may not be necessary when the measures are highly correlated and could decrease the precision of results.^{23 24} If subgroup analyses are of interest to assess heterogeneity of treatment effect, additional interaction terms between treatment arm and the subgroup defining factor should be included in the imputation model to avoid biasing interaction tests towards the null. Alternatively, several have recommended fitting separate imputation models within each randomised group rather than including interaction terms.²⁵

There appears to be little advantage to imputing data at the individual component level and then deriving SRI-5 from the underlying imputed values when patients at specific assessment times generally either have data on all three disease activity measures that comprise SRI or missing data for all of them, that is, no partial information on disease status. Imputing SRI-5 directly as a binary outcome yielded similar results. Others have reached similar conclusions with missing data in composite outcomes used in psychology and rheumatoid arthritis.^{26 27}

Sensitivity analysis should be carried out with different specifications of the imputation model to assess robustness of results to the predictors included in the model.²⁸ Consistency in results across imputation models provides greater confidence in the trial conclusions. Since it is not possible to know the true underlying missing data mechanism and to distinguish MAR from MNAR with the

observed data only, sensitivity analyses should also include assessments of how robust the results are to departures from the MAR assumption. Data that are MNAR will lead to biased results unless the missing data mechanism is known and explicitly modelled in the analysis.

When the missing data rates approaches 50%, results from MI and other missing data methods should be interpreted with caution.¹¹ As a result, strategies to minimise the amount of missing data should be incorporated into the design and conduct of the trial. These strategies include sending patients frequent reminders about their follow-up visits, encouraging participants to remain in the trial, and monitoring the extent of missing data during the trial and taking corrective actions if needed.

Patients in SLE trials who withdraw from treatment because of lack of efficacy or toxicity or need rescue medications not allowed in the protocol are often deemed as treatment failures and no longer followed.^{29 30} This is an example of what the recent International Conference on Harmonisation (ICH) E9 addendum on estimands³¹ refers to as a ‘composite strategy’ for handling treatment withdrawal and other intercurrent events, since the events are integrated into the outcome definition. In an SLE trial that uses this strategy, a successful outcome is not just achieving clinical response by SRI-5 but also ability to tolerate treatment long enough to respond. In this case, additional data following the intercurrent event are not needed nor used if collected, and the resulting treatment effect (estimand) that is estimated is the difference between intervention groups in a combined outcome that requires both disease improvement and treatment adherence. In contrast, if the treatment effect of interest is the ‘treatment policy’ estimand, that is, the effect of the policy of assigning a patient to one treatment versus another as opposed to the treatment effect if taken as directed, then patients who have not withdrawn consent should be assessed for the main outcome, irrespective of treatment adherence. The argument for considering the treatment policy estimand has been that it better reflects the effect of the treatment when used in practice. As emphasised in the ICH E9 addendum, clear specification of the target estimand in the trial is important since this has implications for the collection and handling of data before and after the occurrence of any intercurrent events, and how missing data should be addressed. Having the option to evaluate different estimands in the same study is advantageous since each provides different insights about the treatment effect. We refer the reader to the ICH-E9 report and other papers for further discussion on estimands.

Other missing data approaches, in addition to MI, that would be better alternatives to the single imputation methods commonly used in SLE trials include inverse probability weighting and maximum likelihood, which were not discussed in this paper. Regardless of the missing data method that is adopted, it should be clearly justified and the underlying assumptions on which it is based should be plausible. Of course the best approach

for minimising the effects of missing data is to prevent its occurrence in the first place and obviate the need for applying missing data methods since they often require assumptions that are difficult or not possible to verify.

Contributors MK, JTM, CW, SV, KK and PI contributed to the study design, statistical analysis and interpretation of results. LH contributed to the acquisition of data. All authors reviewed and approved the final manuscript.

Funding This work was supported by a grant from the Lupus Foundation of America.

Competing interests JTM and KK received consulting fees from Eli Lilly. MK received consulting fees from Celgene.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Mimi Kim <http://orcid.org/0000-0001-9651-7124>

REFERENCES

- National Research Council (US) Panel on Handling Missing Data in Clinical Trials. *The prevention and treatment of missing data in clinical trials*. Washington DC: National Academies Press, 2010.
- Stack CB, Butterworth T, Goldin R. Designed learning: missing data in clinical Research Designed learning: missing data in clinical research. *Ann Intern Med* 2018;168:744.
- Isenberg DA, Petri M, Kalunian K, et al. Efficacy and safety of subcutaneous tabalumab in patients with systemic lupus erythematosus: results from ILLUMINATE-1, a 52-week, phase III, multicentre, randomised, double-blind, placebo-controlled study. *Ann Rheum Dis* 2016;75:323–31.
- Merrill JT, van Vollenhoven RF, Buyon JP, et al. Efficacy and safety of subcutaneous tabalumab, a monoclonal antibody to B-cell activating factor, in patients with systemic lupus erythematosus: results from ILLUMINATE-2, a 52-week, phase III, multicentre, randomised, double-blind, placebo-controlled study. *Ann Rheum Dis* 2016;75:332–40.
- Clowse MEB, Wallace DJ, Furie RA, et al. Efficacy and safety of Epratuzumab in moderately to severely active systemic lupus erythematosus: results from two phase III randomized, double-blind, placebo-controlled trials. *Arthritis & Rheumatology* 2017;69:362–75.
- Lachin JM. Fallacies of last observation carried forward analyses. *Clin Trials* 2016;13:161–8.
- Raghunathan TE. What do we do with missing data? some options for analysis of incomplete data. *Annu Rev Public Health* 2004;25:99–117.
- Rubin DB. Multiple imputation after 18+ years. *J Am Stat Assoc* 1996;91:473–89.
- Baron G, Ravaud P, Samson A, et al. Missing data in randomized controlled trials of rheumatoid arthritis with radiographic outcomes: a simulation study. *Arthritis Rheum* 2008;59:25–31.
- Olsen IC, Kvien TK, Uhlig T. Consequences of handling missing data for treatment response in osteoarthritis: a simulation study. *Osteoarthritis and Cartilage* 2012;20:822–8.
- Lee KJ, Carlin JB. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *Am J Epidemiol* 2010;171:624–32.
- White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2011;30:377–99.
- van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Methods Med Res* 2007;16:219–42.
- Azur MJ, Stuart EA, Frangakis C, et al. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res* 2011;20:40–9.
- Spratt M, Carpenter J, Sterne JAC, et al. Strategies for multiple imputation in longitudinal studies. *Am J Epidemiol* 2010;172:478–87.
- Bell ML, Fiero M, Horton NJ, et al. Handling missing data in RCTs; a review of the top medical journals. *BMC Med Res Methodol* 2014;14.
- Kenward MG, Carpenter J. Multiple imputation: current perspectives. *Stat Methods Med Res* 2007;16:199–218.
- Moons KGM, Donders RART, Stijnen T, et al. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol* 2006;59:1092–101.
- Rubin D. *Multiple imputation for nonresponse in surveys*. New York: John Wiley, 1987.
- Kim M, Merrill JT, Kalunian K, et al. Evaluating duration of response to treatment in systemic lupus erythematosus clinical trials. *Lupus Science & Medicine* 2018;5:e000266.
- Little R. Pattern-Mixture models for multivariate incomplete data. *J Am Stat Assoc* 1993;88:125–34.
- Ratitch B, O'Kelly M, Tosiello R. Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. *Pharm Stat* 2013;12:337–47.
- Lee KJ, Roberts G, Doyle LW, et al. Multiple imputation for missing data in a longitudinal cohort study: a tutorial based on a detailed case study involving imputation of missing outcome data. *Int J Soc Res Methodol* 2016;19:575–91.
- Hardt J, Herke M, Leonhart R. Auxiliary variables in multiple imputation in regression with missing X: a warning against including too many in small sample research. *BMC Med Res Methodol* 2012;12:184.
- Sullivan TR, White IR, Salter AB, et al. Should multiple imputation be the method of choice for handling missing data in randomized trials? *Stat Methods Med Res* 2018;27:2610–26.
- O'Keeffe AG, Farewell DM, Tom BDM, et al. Multiple imputation of missing composite outcomes in longitudinal data. *Stat Biosci* 2016;8:310–32.
- Graham JW. Missing data analysis: making it work in the real world. *Annu Rev Psychol* 2009;60:549–76.
- White IR, Carpenter J, Horton NJ. Including all individuals is not enough: lessons for intention-to-treat analysis. *Clin Trials* 2012;9:396–407.
- Navarra SV, Guzmán RM, Gallacher AE, et al. Efficacy and safety of belimumab in patients with active systemic lupus erythematosus: a randomised, placebo-controlled, phase 3 trial. *The Lancet* 2011;377:721–31.
- Furie R, Petri M, Zamani O, et al. A phase III, randomized, placebo-controlled study of belimumab, a monoclonal antibody that inhibits B lymphocyte stimulator, in patients with systemic lupus erythematosus. *Arthritis & Rheumatism* 2011;63:3918–30.
- International Conference on harmonisation of technical requirements for pharmaceuticals for human use: ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials on the Guideline on statistical principles for clinical trials 2017.