**Appendix. Technical details for reference-based imputation**

*Multiple imputation-based approaches*

Let

n = the number of patients.

q = the number of treatments

t = the number of visits (assessment times)

i = 1…n the individual patient indicator

J = 1…q the treatment randomly assigned to the ith patient

k = 1…t the individual visit indicator

r = 1…q the reference treatment arm

p = the visit at which patient i withdraws

s = an index for the covariates

The basic underlying model is the standard Gaussian repeated-measures model for quantitative data with parameters for treatment-by-time interaction and for all the baseline covariates.

The patients are independent and the vector for the i’th patient has a multivariate Normal distribution with covariance matrix **.** If the data were complete, then this repeated-measures model could be fitted with the following features.

* A treatment-by-time interaction to indicate the profile across time within each treatment arm, .
* Other covariates, which may or may not be crossed with time, are defined in the matrix . .
* An unstructured covariance matrix which may be shared across arms or may be separate for each arm.

The same model can be fit when some Y are missing by assuming MAR. Estimated parameters from this model are used to build possible predicted profiles for the unobserved values in a separate imputation model, one for each pattern of withdrawal. These profiles represent outcomes to expect in patients after withdrawal. Missing values for an individual are imputed based on the profile of predicted means that matches their treatment and their time of withdrawal.

Three models are required in this process.

1. A parameter-estimation model, which is effectively the MMRM model described above and is fitted assuming MAR. This is a marginal model, not a pattern-specific model.
2. The imputation model, in which for each pattern of withdrawal we build a model for predicted values using the parameters obtained from the parameter estimation model (1). How the parameters in these models link to the parameters estimated in (1) define the different reference-based methods.
3. An analysis model, which analyses the complete data once for each draw (imputation). This is often simply a univariate ANOVA model.

The crucial MNAR part of the process is how the profile across time is defined for each pattern and treatment in the imputation model, based on the values of the parameters in the estimation model. The only difference between patterns is in the parameters of the treatment-by-time interaction, which are the profiles across time for each treatment arm. In the estimation model there is a single set of q×t parameters , while in the imputation model there is a set of q×t parameters, one for each possible withdrawal visit (p=0...t) where apart from the last visit p=t, the data are missing **after** the pth visit. When p=0 all data for the patient are missing.

Different MNAR models are defined by specifying the relationship between and . For any withdrawal pattern p, the response vector **Y** splits into two parts: the observed values **Y**1 prior to withdrawal and the unobserved values **Y**2 which we need to impute. For a single imputation (drawn from the posterior distribution of the parameters), the imputation model has a mean with two parts **μ**1 and **μ**2 and a variance-covariance matrix which splits into four components **Σ**11, **Σ**12, **Σ**21, which is the transpose of **Σ**12 and finally **Σ**22.

Imputed values are sampled by drawing from the distribution of **Y**2 conditional upon the observed value of **Y**1. This distribution is also multivariate Normal with mean **μ**2 **+ Σ**21 **Σ**11-1(**Y**1 **- μ**1) and variance-covariance matrix **Σ**22  - **Σ**21 **Σ**11-1 **Σ**12. In the model definition, acts as an intercept for each combination of treatment and time. Its value is thus dependent upon the constraints applied in the definition of the covariate design matrix Xi. The patient’s individual covariate offset is the same whatever treatment j they receive but varies across times k. We calculate the average of these profiles across all the patients and re-express the estimation model as

[9]

and the imputation model as

In this way is invariant under change to the constraints. In other words, is defined as the treatment-by-time intercept for a patient with average profile of covariate offset. It is the offset for an “average” patient.

The different methods are defined in terms of these average set of intercepts.

* **MAR Missing At Random**

for all patterns p, j and k. [11]

Here all the patterns have the same profile, so that, conditional on covariates, missingness is independent of the profile. This is equivalent to an MMRM model and estimates a *de jure* estimand.

* **CIR Copy Increment from Reference**

= for k≤p

= for k>p [12]

Here all the patterns have the same profile up to withdrawal (that estimated for this arm in the parameter estimation model), but after withdrawal the pattern tracks parallel to the pattern for the reference arm. For the reference arm this is equivalent to MAR. The assumption is that when a patient stops treatment they continue to take advantage of their previous therapy, but from withdrawal onwards they progress in the same way as the patients in the reference arm (parallel to the profile for the reference arm). As such, it will usually answer a *de facto* question and be especially applicable for drugs that modify the progression of disease.

* **Jump to Reference**

= for k≤p

= for k>p [13]

Here all the patterns have the same profile up to withdrawal (that estimated in the parameter estimation model), but after withdrawal the profile jumps to the estimated profile for the reference arm. As for CIR the reference arm profile is equivalent to MAR.

* **Copy Reference**

= for all p, j and k. [14]

Here all patterns, except patterns for those who do not withdraw, follow the whole profile estimated for the reference arm in the parameter estimation model, both before and after withdrawal. As a result, this method is distinctly different from the rest of the methods in that the profile prior to withdrawal is that for the reference arm rather than that for the patients own arm. Hence the residuals (**Y**1 **- μ**1) are measured from mean **μ**1 for the reference arm, rather than that for the patients own arm in the formula **μ**2 **+ Σ**21 **Σ**11-1(**Y**1 **- μ**1) for the conditional mean used in imputation. As for CIR and J2R, the reference arm profile for CR is equivalent to MAR.

*Reference-based approaches using maximum likelihood*

Using the same notations as above, the MMRM model defines the mean responses for patient at time

[15]

where is the indicator for treatment group, is a collection of covariates to be adjusted in the analysis model which are centralized as in [9], is the vector of corresponding coefficients, and is the treatment effect at time . We also assume that the response vector has mean as defined in [15] and an unstructured covariance matrix , and the coefficients and covariance matrix are the same across treatment groups. With centralized , the can be interpreted as the treatment effect at the mean for group at time . Hence the treatment difference at time can be constructed from the parameters of . For example, in a trial with two treatment groups, say test drug (k=1) and placebo (k=2), then the treatment difference at last time point is defined as . We use the superscript MAR to indicate that this treatment difference is for a *de jure* estimand under MAR, which assesses the mean difference between patients in group 1 and that in group 2 assuming they stay in the study and take the assigned treatments over the time points.

When there are no missing data, this treatment difference can be estimated from the MMRM or even a univariate ANCOVA model for the observations with factors of treatment and baseline covariates. The results from this univariate ANCOVA model will be the same as that from the MMRM because no restriction is applied to the covariance matrix or mean profile across time points in the MMRM analysis. When there are missing data, the parameters can be estimated from the MMRM which implicitly assumed following a multivariate normal distribution. Under MAR, the parameters in [15] and the treatment differences can be estimated using likelihood constructed from all observed data (the missing data can be ignored). It addresses a hypothetical estimand of what would be happen when all patients follow the study protocol and complete the study up to time

The reference-based controlled imputation addresses a different estimand which is the difference between treatments regardless of adherence and without initiation of rescue medication. As described in the previous section, the mean profile for response is defined as

[16]

where is the last time point with observation for patient . So are missing. The values are specified based on the reference-based controlled imputation strategy.

Assumptions are made for the mean parameters under the reference-based controlled imputation strategy. For reference group (e.g., placebo), it is assumed that across time points, that is, the means for patients who dropped out from reference group and without taking any rescue medication are the same to those who stay in the study and taking placebo. This is reasonable in clinical trials where no psychologic effects are anticipated, so no treatment is like taking placebo. For the other treatment groups, the means of a patient who dropped out at time are based on the assumed strategy using the reference group. Specifically,

[17]

To simplify the notations, we ignore the patient indicator, and let be the observed sub-vector for this patient, and be the sub-vector for missing data, and

be split sub-vectors and block matrices with dimensions corresponding to the and . Then the missing data vector will be imputed from a conditional distribution

.

Hence the unconditional mean sub-vector for the missing data (evaluated at ) after applying the reference-based controlled imputation is

[18]

where . For completers, we have |.

Suppose be the proportion of patients who have data in group at time , and let be the proportion of completers for group Then the overall mean at last time point for treatment can be expressed as

[19]

where is the last element (at time t) of the sub-vector. Therefore, the treatment effect under reference-based imputation can be expressed as a linear combination of the parameters of in [15] and the proportions of patients in each missing data pattern (i.e., time of dropout).

The treatment difference between groups can be constructed from the overall means as given in [19]. In general, it will be a function of , and , . One of the treatment groups, commonly the placebo group in placebo-controlled trials, will be used as the reference group. The treatment difference between group and reference group is . Here we use the superscript RBI to indicate for reference-based imputation approach.

With the expression of [19], the can be estimated using likelihood approach. Specifically, the parameters of can be obtained from the MMRM using likelihood method. Suppose is the observed proportions of dropouts over time or completers, are estimated from the MMRM analysis. Then the point estimate and its standard error for a treatment difference can be derived. First, we treat as fixed. Hence the treatment difference will be estimated as a linear combination of .

[20]

To obtain its variance, we use the formula

. [21]

The conditional variance of the first term given can be derived from the variance estimates from MMRM model (for example, using the covariance estimates of the LSMEAN differences from the SAS PROC MIXED analysis). The second term can be calculated using the point estimates of and , with for i=k, for , where is the sample size for group . As an example, let us derive the variance for J2R in a trial with two treatment groups, say test drug (k=1) and placebo (k=2). Hence the treatment difference at last time point with J2R is defined as , where is the proportion of completers in the test drug group. Suppose be the MMRM estimate for the treatment difference with variance estimate . Then the treatment difference under J2R will be estimated as and its variance estimate is

. [22]

It should be noted that the estimates and inference for RBI parameters are all based on the parameters in [15], that can be estimated using MMRM under MAR assumption. The MAR assumption here is used to define the latent mean profiles of [15]. The treatment difference as constructed from contrasting among is the estimand of interest, which is based on a special pattern mixture model as specified by the RBI strategy. It is assumed that the missing data for the reference group are MAR. However, the distribution of values after withdrawal is different from those who stay in the study for the non-reference groups. Therefore, the missing data mechanism for RBI is missing not at random (MNAR).

We should also distinguish that the MNAR approach for RBI is different from assuming MNAR for equation [15] to allow missing data depending on the unobserved data. The later will cause the parameters of undefined based on observed data without knowing the missing data mechanism. So if the “true” parameters (that we never know in real applications) in [15] are from MNAR, the RBI analysis will still use MAR to define the latent parameters for [15] based on all observed data. These latent parameters will be different from the “true” parameters under MNAR. Therefore, under this MNAR scenarios, the RBI analysis will be biased in estimates and may inflate type-I error as shown in simulations by Liu and Pang (2016). The RBI methods are special pattern mixture models under the principle of reference-based controlled imputation that are defined using the latent parameters obtained under MAR.

*Reference-based approaches using Bayesian methods*

Bayesian methods for context of reference-based imputation have been proposed (Liu and Peng, 2016). With advancements in computation and statistical software, simulation-based methods such as Markov Chain Monte Carlo (MCMC) sampling become a feasible alternative for analysis of longitudinal clinical trials. Computationally, MCMC generates a series samples for model parameters to approximate their equilibrium posterior distributions, and then inference is made based on these distributions. Bayesian methods start with specification of prior distributions and initial values for model parameters. In applications with no specific prior information, non-informative or flat priors are used, and statistical inference is similar to the likelihood-based methods. The samples are updated iteratively using direct sampling when explicit posterior distributions are available such as under conjugacy of priors or using data augmentation methods such as Metropolis or Metropolis-Hastings algorithms for complicated distributions (Tanner and Wong, 1987).

Consider how the Bayesian methods may be used to fit a repeated measures model in a longitudinal study with no missing data. Assume, where mean vector is defined in [15] with parameters of and . The likelihood function can be written as

where is the density function of a multivariate normal with mean and covariance matrix . Suppose is the prior for parameters in the mean , and is a prior for covariance. Then the posterior distribution for will be

[23]

Without specific prior information for the parameters, the non-informative prior may be used. For example, , and an inverse Wishart distribution with dimension , and is an identity matrix of the same dimension. Then the posterior distribution samples can be obtained through MCMC process for [23] using many Bayesian computational software packages such as SAS PROC MCMC and STAN. The posterior samples for any function of these parameters, e.g., the treatment difference between groups at last time point , , can be obtained. Statistical estimates such as mean and standard deviation and credible intervals can be obtained from these posterior samples.

In longitudinal studies with missing data, the missing values are treated as parameters in Bayesian analysis and directly sampled from the conditional distribution. For a repeated measures model with an MAR missing data assumption, simplify the notations by ignoring the patient indicator, and let be the observed sub-vector for a patient, and be the sub-vector for missing data, and

are sub-vectors and block matrices with dimensions corresponding to the and . Using the Bayesian approach, the missing data are directly sampled from

. [24]

Then the samples of model parameters can be drawn from the posterior distribution [23] based on combined observed data and sampled missing data . Starting with the giving initial values of the parameters, the sampling process is carried out iteratively as follows,

1. Sampling as in [24] across patients in the study,
2. Sampling as in [23],
3. Calculating parameter of interest, e.g., .

Statistical estimates such as mean and standard deviation and credible intervals for can be obtained from the posterior samples for statistical inference.

To reduce the impact of initial values on the posterior distribution samples, it is common practice to discard a portion of the MCMC samples (called burn-in) and use the samples after the burn-in for statistical inference. For complicated distributions, the mixing (or convergence) of the MCMC samples may be slow. It is therefore important to check the convergence of MCMC samples for all parameters through, e.g. visually reviewing trace plots, or checking some test statistics as provided in Bayesian computation software, e.g., SAS PROC MCMC (SAS/STAT 14.3).

It should be noted that the Bayesian approach for the repeated measures analysis also requires MAR or ignorable missing data assumption. Specifically, the direct sampling for the missing data [24] is based on the same conditional distribution as those who complete the study. With non-informative prior distributions, the inference from Bayesian analyses would be very similar to that from the likelihood-based method using mixed models. Simulations by Liu and Pang (2016) showed that the point estimates and standard deviations from the MCMC approach for the treatment difference were like that obtained from the mixed model analysis, so were the statistical inference such as type-I error rates and power for testing the treatment difference.

For reference-based imputation approaches, the treatment difference can be expressed as a function of and as given in [14]. Bayesian analysis will treat all these unknown parameters as random variables. To reflect the randomness of the proportions of missing data pattern, we may sample the proportion of missing data pattern from a Dirichlet distribution with probability density

[25]

where >0 is the proportion of dropout for treatment group at time or completers for ; ;and is the observed number of patients at treatment group who dropped out at time or completers for . Specifically, starting with the initial values of the parameters, the sampling process is carried out iteratively as follows,

1. Sampling as in [24] across patients in the study,
2. Sampling as in [23],
3. Sampling as in [25],
4. Calculating parameter of interest, e.g., .

Statistical estimates such as mean and standard deviation and credible intervals for can be obtained from the posterior samples for statistical inference.

Simulations by Liu and Pang (2016) also showed that the point estimates and standard deviations from the posterior samples for treatment difference under reference-based imputation strategies were very similar to that obtained from approximation using likelihood approach as described in formulas [15] and [16], so were the statistical inference such as type-I error rates and power for testing the treatment difference. In this Bayesian approach, the variation of the dropout proportion is incorporated into the posterior distribution through sampling the from the Dirichlet distribution.

*Appendix 4.4. Considerations for the variance of reference-based estimators*

Important considerations are embedded within the computations for the variance of the various reference-based estimators.  In reference-based methods using multiple imputation, the variance and standard error of treatment contrasts are determined using the same method as in MAR-based multiple imputation.  Rubin’s rules are used to combine the between and within imputation variances.  This approach explicitly imputes missing values for each patient in the experimental group using a model developed from the reference group.  It explicitly accounts for the uncertainty (variability) in the imputed values when determining the variance and standard error of treatment contrasts, while maintaining the statistical independence of the reference and experimental arms.  However, this MI-based variance may over-estimate the true variability as compared to the empirical variance from simulations12 38.

The Likelihood- and Bayesian-based approaches to group mean reference-based imputation presented above take a different approach.  Rather than imputing individual values with uncertainty, these methods define the mean of patients in the experimental group who discontinued to be equal to the mean of the control group.  Because imputation is done by definition, the variance and standard errors of treatment contrasts reflect no uncertainty in the imputation. Therefore, group mean based approaches generally yield smaller standard errors than the corresponding MI-based methods.  In fact, the standard errors will also be smaller than in the corresponding MMRM analysis.  Intuitively, the standard error when imputing data should not be smaller than the standard error based on the observed data.  In fact, when considering equation 15, one sees that if every patient dropped out, the point estimate and the variance would both be zero.  However, the approach used to obtain the variance of treatment contrasts from group mean based methods does not require explicit data imputation andcan be justified using the assumptions of reference-based imputation*.*

Specifically, if we define that patients who discontinue an experimental drug do not benefit from it, we know their outcome – zero benefit. Group mean imputation uses the mean of the reference group to ascribe zero benefit to experimental group dropouts.  This approach of defining the outcome is consistent with composite approaches to dealing with ICEs wherein dropout is considered an outcome – here that outcome is zero benefit.  This variance shrinkage is also seen in other methods for obtaining the variance and standard error of treatment contrasts, such as bootstrapping.  Simulation studies indicate that the average of these likelihood- and Bayesian-based variance estimates is close to the empirical variance from simulations 19, 38.

Alternative approaches to determining the variance for group mean reference-based imputation can incorporate uncertainty of imputation.  Each approach has its own assumptions and conceptual considerations. The point here is not to lobby for one approach or another.  Rather, the intent is to make clear that practical and conceptual differences exist.  Therefore, the estimator – and the method for assessing the variance of the estimate - should match the estimand.