Statistica Sinica Preprint No: SS-2016-0332	
Title	Comment on "Dissecting Multiple Imputation from a Multi-phase Inference Perspective: What Happens When God's Imputer's and Analyst's Models are Uncongenial?" by X.Xie and X.Meng.
Manuscript ID	SS-2016-0332
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0332
Complete List of Authors	Roderick Little and Tingting Zhou
Corresponding Author	Roderick Little
E-mail	rlittle@umich.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0332

Statistica Sinica

Discussion

Roderick Little and Tingting Zhou

University of Michigan

Xie and Meng's paper is a theoretical tour de force, providing further insight into the performance of multiple imputation combining rules when the imputer and analysis models differ. Implications for practice are not entirely clear, at least to us; one conclusion is to continue to use the MI combining rules, while seeking to minimize differences between the imputer and analyst models, or attempting to ensure that the differences are in the direction of making the MI combining rule conservative. Another conclusion is to abandon the Rubin's rules in favor of Xie and Meng's more conservative ones, although the penalties in increased width of confidence intervals seem stiff. The choice is an example of a basic question that applies to all statistics, namely what aspects of potential model misspecification should be formally reflected in measures of uncertainty. Xie and Meng's examples are instructive but perhaps not very realistic, and we describe here an extension of Example 1 that is very relevant to an applied setting.

An area where multiple imputation is receiving increased attention is in handling missing data in clinical trials. A National Research Council study (National Research Council (2010), Little et al. (2012)) advocates sensitivity analysis as an important component of the analysis of clinical trial data, and since that report there has been much activity to develop new methods

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0332

and software (e.g. Mallinckrodt, Lin, and Molenberghs (2013), Ratitch, O'Kelly, and Tosiello (2013), Liublinska and Rubin (2014), Little et al. (2016)). The tricky modeling problem is to decide the appropriate range of models to consider in such an analysis: a narrow class may miss important possible scenarios, whereas a broad class that includes implausible models, such as worst case scenarios where dropouts are all considered treatment failures in the treated group and treatment successes in the control group, leads to excessively high ranges of uncertainty.

A convenient approach to sensitivity analysis, which is relatively easy to implement and convey to clinicians, models departures from missing at random via one or more sensitivity parameters that characterize differences between participants who do and do not drop out in each treatment group, after controlling for observed characteristics. This approach leads naturally to pattern mixture models (Little (1993)) where distributions of trial outcomes are modeled conditional on the dropout indicator. Formally let D be a variable with value 1 for dropouts and 0 for participants who do not drop out. The joint distribution of D and trial outcomes Y is factored as:

$$f_{Y,D}(Y, D|Z, X, \phi, \delta) = f_{Y|D}(Y|D, Z, X, \phi, \delta) f_D(D|Z, X, \pi) \tag{1}$$

where Z is a treatment indicator, and X represents other fully observed covariates. More generally, D may have more than two values, corresponding to different drop-out times. The sensitivity analysis involves varying sensitivity parameters δ, a low (one or two-) dimensional parameter that characterize differences between $f_{Y|D}(Y|D = 0, Z, X, \phi, \delta)$ and $f_{Y|D}(Y|D = 1, Z, X, \phi, \delta)$; δ is generally not identified from the data, so the sensitivity analysis assesses the treatment effect over a range of plausible values of δ, or the size of δ is computed and assessed at the tipping point where statistical significance of the treatment effect is lost.

A practical approach to implementing this sensitivity analysis is to multiply-impute values of Y after dropout for each preset value of δ, and provide inferences for the parameters

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0332

characterizing the treatment effect using Rubin's MI combining rules. This leads to potential uncongeniality, since the natural analysis model for Y is a model $f_Y(Y|Z, X, \theta)$ for the distribution of Y given Z, X in the absence of missing data; this analysis model is often incompatible with the imputation model of form (1) . If the imputation model does not correspond to the model that generated data, resulting inferences clearly have the potential for bias. A more subtle question is the validity of MI inferences based on the model $f_Y(Y|Z, X, \theta)$ when imputations are generated under the correct model that generated the data. To shed light on this issue, we describe the results of a small simulation study, based on a realistic extension of the Xie and Meng's Example 1.

Repeated univariate samples of size $N = 50$ for an outcome Y and drop-out indicator D are generated from a simple version of (1), with Z and X null, $\phi = (\mu_0, \sigma^2)$, and $\delta = (\delta_1, \delta_2)$:

$$D \sim Bernoulli(\pi)$$

$$\begin{aligned} Y|D = 0 &\sim N(\mu_0, \sigma^2) \\ Y|D = 1 &\sim N(\mu_0 + \delta_1, \delta_2^2\sigma^2) \end{aligned} \tag{2}$$

The resulting missing-data mechanism is missing not at random unless $\delta_1 = 0, \delta_2 = 1$; the sensitivity parameters are $\delta = (\delta_1, \delta_2)$, which model differences in the mean and variance of the distribution of Y for respondents and drop-outs. The marginal distribution of Y based on (2) is a mixture of normals, with mean $\theta = \mu_0 + \pi\delta_1$ and variance $\tau^2 = (1-\pi)\sigma^2 + \pi\sigma^2\delta_2^2 + \pi(1-\pi)\delta_1^2$. The target parameter is the overall population mean θ. A sample thus has n respondents with Y measured and $N-n$ dropouts with Y missing, where n is Binomial with index N and probability π.

Missing values of Y were multiply imputed (with 100 imputations) using their posterior

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0332

predictive distribution, based on the correct model (2) that generated the data, assuming in particular the correct choice of sensitivity parameters δ, with Jeffreys' prior distributions for the parameters ϕ. The resulting MI data sets were analyzed using Rubin's combining rules, for two choices of analysis models:

(a) the pattern-mixture model that generated the data, again with the correct choice of δ, and

(b) the standard univariate normal model for the complete data,

$$Y \sim N(\theta, \tau^2) \tag{3}$$

Tables 1 and 2 show empirical bias, bias, root mean squared error, and 95% confidence interval coverage for the two MI analyses, over 1000 replicate data sets. In Table 1, we set $\delta_2 = 1$ and varied δ_1 from 0 to 3, thus varying the difference in means for respondents and dropouts. In Table 2, we set $\delta_1 = 0$ and varied δ_2 from 0.2 to 5, thus varying the differences in variances for respondents and dropouts.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0332

Table 1: Empirical Bias*1000, Root Mean Squared Error *1000 and Confidence Interval Noncoverage (Nominal = 50) over 1000 simulated data sets of sample size of 50, for MI Inferences Under (a) PMM = the Pattern-Mixture model (2) that generated the data, with correct choice of δ, and (b) NOR = the normal complete-data model (3). $\delta_2 = 1$ and δ_1 varied from 0 to 3.

	δ_1					
$\delta_2 = 1$	0	0.5	1.0	1.5	2.0	3.0
Bias PMM	-2	0	2	4	6	10
Bias NOR	-1	-1	-1	-1	-1	-1
RMSE PMM	251	252	252	252	252	253
RMSE NOR	252	252	252	252	252	252
Noncov PMM	48	45	40	30	23	9
Noncov NOR	43	40	32	23	21	6

.

In both sets of simulations in Tables 1 and 2, Bayes inference based on the pattern-mixture model that generated the data had small empirical bias, confidence coverage that was close to nominal or conservative. Bayes inference for the normal model yielded small empirical bias, and confidence coverage close to nominal or conservative in Table 1, where the mean is being varied but the variance is held constant. However in Table 2, the MI inference for the normal model had close to nominal coverage when $\delta_2 = 1$, conservative coverage when δ_2 was much less than

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0332

Table 2: Empirical Bias*1000, Root Mean Squared Error *1000 and Confidence Interval Noncoverage (Nominal = 50) over 1000 simulated data sets of sample size of 50, for MI Inferences Under (a) PMM = the Pattern-Mixture model (2) that generated the data, with correct choice of δ, and (b) NOR = the normal complete-data model (3). $\delta_1 = 0$ and δ_2 varied from 0.2 to 5.

	δ_2				
$\delta_1 = 0$	0.2	0.5	1	2	5
Bias PMM	-2	-2	-2	-2	-2
Bias NOR	-2	-2	-1	-1	0
RMSE PMM	251	251	251	251	251
RMSE NOR	252	252	252	253	260
Noncov PMM	48	48	48	48	48
Noncov NOR	115	98	43	1	0

.

1, and anti-conservative when δ_2 was much greater than 1. These results are consistent with the results in Example 1 of Xie and Meng, and suggest that analyses under the normal model are robust to sensitivity analyses that concern deviations in the means between respondents and nonrespondents, but are less robust to sensitivity analyses that concern deviations in the variances. Basing the inference on the pattern-mixture models that generated the imputations yields more coherent results, although it deviates from current practice.

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0332

References

Little, R.J. (1993). Pattern Mixture Models for Multivariate Incomplete Data. *J. Amer.Statist. Assoc.*, **88**, 125-134.

Little, R.J., D'Agostino, R., Cohen, M.L., Dickersin, K., Emerson, S.S., Farrar, J.T., Frangakis, C., Hogan, J.W., Molenberghs, G., Murphy, S.A., Rotnitsky, A., Scharfstein, D., Neaton, J.D., Shih, W., Siegel, J.P., and Stern, H. (2012). Special Report: The Prevention and Treatment of Missing Data in Clinical Trials. *New England J. Medicine*, **367**, 1355-1360.

Little, R.J., Wang, J., Sun, X., Tian, H., Suh, E-Y., Lee, M., Sarich, T., Oppenheimer, L., Plotnikov, A., Wittes, J., Cook-Bruns, N., Burton, P., Gibson, M., and Mohanty, S. (2016). The Treatment of Missing Data in a Large Cardiovascular Clinical Outcomes Study. *Clin. Trials* **13**, 344-351.

Liublinska V. and Rubin, D.B. (2014). Sensitivity Analysis for a Partially Missing Binary Outcome in a Two-Arm Randomized Clinical Trial. *Statist. Med.* **33**, 4170-4185.

Mallinckrodt, C.H., Lin, Q. and Molenberghs, M. (2013). A structured framework for assessing sensitivity to missing data assumptions in longitudinal clinical trials. *Pharm Stat. 2013*; **12(1)**:1-6.

National Research Council (2010) The Prevention and Treatment of Missing Data in Clinical

Statistica Sinica: Preprint
doi:10.5705/ss.202016.0332

Trials. *National Academy Press: Washington DC.*

Ratitch, B., OKelly, M. and Tosiello, R. (2013). Missing Data in Clinical Trials: from Clinical Assumptions to Statistical Analysis Using Pattern-Mixture Models. *Pharmaceut. Statist.* **12**, 337-347.

University of Michigan

E-mail: (rlittle@umich.edu)

E-mail: (tkzhou@umich.edu)