

Assessing Approaches for Handling Missing Measures of Precision in Model-Based Meta Analysis (MBMA)

Claire Kong^{1*}, John Maringwa², Chandni Valiathan¹

¹ Clinical Pharmacology & Pharmacometrics, Johnson & Johnson Innovative Medicine, USA

² Clinical Pharmacology & Pharmacometrics, Johnson & Johnson Innovative Medicine, Netherlands

ABSTRACT

This project investigated several methods for handling missing measures of precision in Model-Based Meta Analysis (MBMA). Using data from 16 placebo-controlled studies on anti-depression treatments (venlafaxine and fluoxetine), 6 approaches to address the missingness issue were explored: (1) weighting outcomes by sample size (N), (2 & 3) replacing all or only missing values based on the weighted mean of available standard deviations (SDs), (4 & 5) replacing all or only missing values with the maximum reported SD (reflecting the "worst-case" scenario), and (6) imputing missing values using random forest (RF). To establish a baseline for comparison, estimates based on RF approach were regarded as the "true" values for simulating data. Next, permutations of % missing precision values, sample size and approach created a "scenario", and 1000 replicates were simulated per each "scenario".

Results showed that imputing missing SDs by weighted mean and by RF consistently had the lowest relative standard error (%RSE) across levels of missing variance. Moreover, when around 70% of the SDs were missing, the %RSE associated with replacing missing SDs by the maximum became comparable to that of weighting by N. This suggests that the pragmatic approach of weighting by sample size may be conservative, leading to wider confidence intervals. Thus, when some measures of precision are reported, it may be advisable to impute missing values using the weighted mean of available information or by RF.

BACKGROUND

- Model-based meta-analysis (MBMA) is a method that uses published literature data to develop models that help address clinical design questions.
- A typical continuous outcome reported literature aggregate level data point in MBMA consists of a mean value and an associated measure of precision (standard error (SE), or SD).
- When reported SEs are used as fixed weights in the model, no further estimation of residual variance is required (technically, "fixing sigma to 1").
- However, measures of precision are often unreported in literature sources, creating a missingness issue, but they are necessary weights in an MBMA model.

OBJECTIVE

- Use simulations to explore the impact of different extents of missing precision measures in MBMA and assess various approaches to address missingness issue .
- Provide recommendations on which approach to use, depending on the extent of missing information in terms of precision measures.

METHODS

Motivating Example

- Published literature data from 16 placebo-controlled trials of 2 anti-depression treatments for major depressive disorder (MDD) was used
- Venlafaxine (10 trials, 1289 pts, doses ranging from 25 to 375 mg/day), fluoxetine (8 trials, 982 pts, doses from 20, 40, to 60 mg/day), and 1161 placebo-treated pts
- Change from baseline in the Hamilton Depression Rating (HAMD) scale at primary timepoint was the clinical endpoint of interest

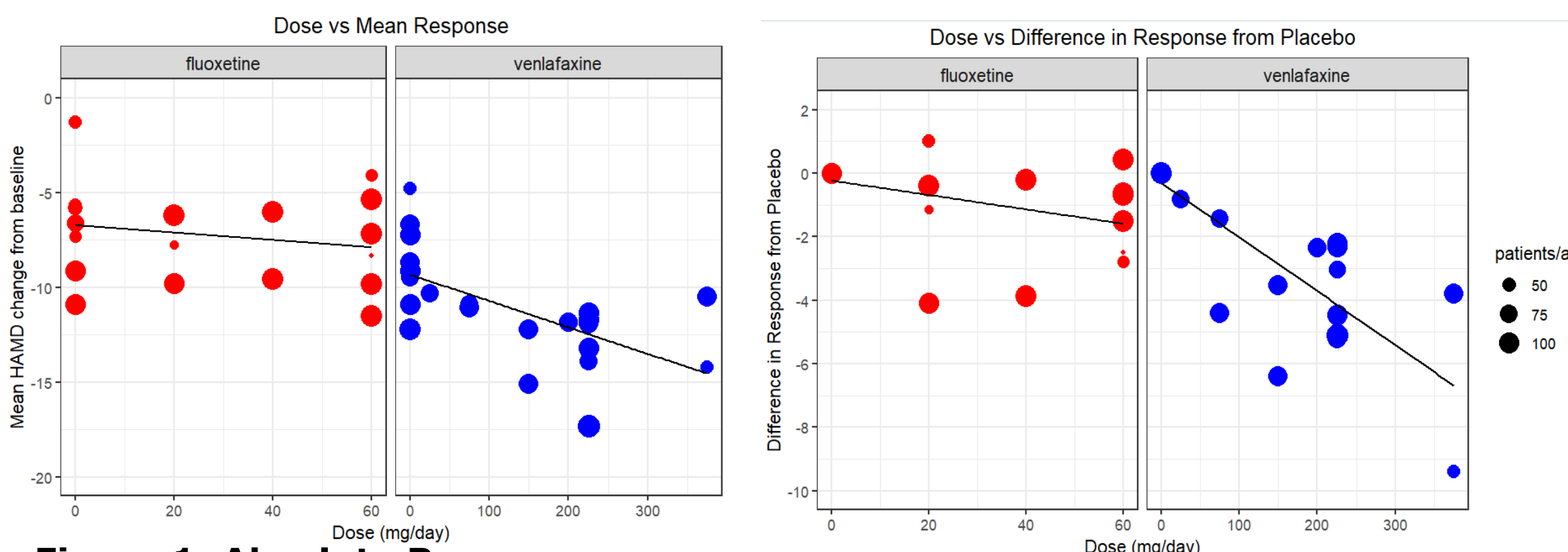


Figure 1. Absolute Response

Venlafaxine changes appear larger, with a hint of a dose-dependent relationship

Figure 2. Treatment Effect

Venlafaxine treatment effect (placebo-corrected) appears more pronounced

Model Development

Based on exploratory data analysis, several model structural forms were investigated (see Table 1). Likelihood ratio tests were performed on nested models to assess drug effect (Table 2).

Model	Description	DF	Log-Lik	AIC	Nested Models	Deviance Difference	P-value
1	Placebo effect only	16	-110.9	253.8			
2	Common effect for fluoxetine and venlafaxine	17	-68.5	171.1	Model 1 vs. Model 2	84.7	<0.001
3	Overall fluoxetine and overall venlafaxine	18	-56.4	148.8	Model 1 vs. Model 3	109.1	<0.001
4	Linear fluoxetine and linear venlafaxine	18	-64.2	164.4	Model 2 vs. Model 3	24.3	<0.001
5	Linear fluoxetine and Emax venlafaxine	19	-57.2	152.3			

Table 2. LRT of Models 1, 2, 3

Table 1. Candidate Model Evaluations

- Model 3 had the lowest AIC, but based on graphical fit and to facilitate simulations showing dose impact (including non-linear trends), Model 5 was selected for further consideration
- Deviance difference is highest for Model 1 vs. Model 3 - indicative of drug effect but not of dose-dependency

Model Formulation

Linear fluoxetine & Emax venlafaxine

$$Y_{ij} = \begin{cases} eo_i + slp_k * dose_{ijk} + \varepsilon_{ij}, & k = \text{fluoxetine} \\ eo_i + \frac{Emax_k * dose_{ijk}}{ED50_k + dose_{ijk}} + \varepsilon_{ij}, & k = \text{venlafaxine} \end{cases}$$

Y_{ij} : mean HAMD change in arm j of study i

eo_i : non-parametric, study-specific placebo response

slp_k : fluoxetine slope

$Emax_k$: venlafaxine maximal effect with potency $ED50_k$

$dose_{ijk}$: dose of drug k in arm j of study i

ε_{ij} : residual error terms

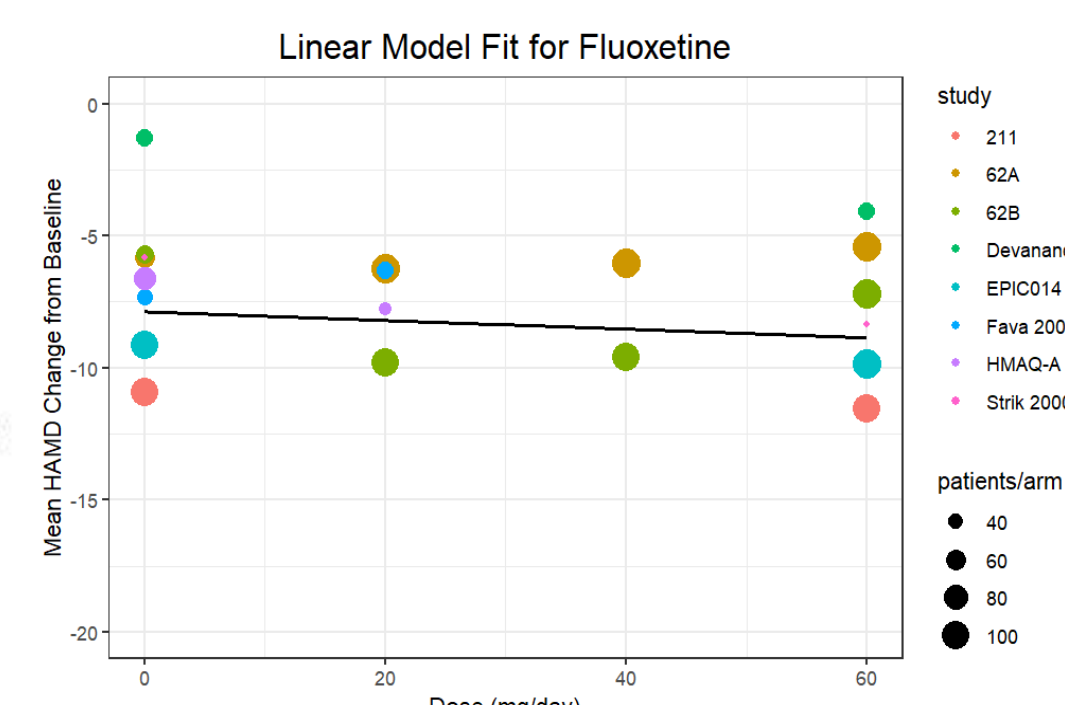


Figure 3. Final Model Fit

Simulation Settings

Setting	Observed data	Simulation
# Simulations	16	1000 per scenario*
Trials	16	20
Dose Levels	Fluoxetine: 0, 20, 40, 60 mg/day Venlafaxine: 0, 25, 75, 150, 200, 225, 375 mg/day	Same
# Patients/Arm (N)	Range: 27-114	50, 100, or 150
% Missing Precision	49%	5%, 10%, 20%, 50%, 70%, 90%
Missing Precision Imputation Method	Methods applied: 1. Weighting by sample size N 2. Replace all SDs by weighted mean of available SDs 3. Replace only missing SDs by mean of available SDs 4. Replace all SDs by maximum of available** SDs 5. Replace only missing SDs by maximum of available** SDs 6. Impute missing SDs by random forest	

* Scenario: combination of N, missing % variance, and imputation method

** Replacing by the maximum of available SDs reflects the "worst-case scenario" in terms of precision values

Table 3.

- Parameter estimates from the RF imputed model on observed data were considered as "true" values
- Model uncertainty incorporated by resampling 1000 times from the variance-covariance matrix
- Between-subject variability based on resampling SDs from observed data with RF imputation

RESULTS

	Fluoxetine	Venlafaxine	
	Slope (%RSE)	Emax (%RSE)	Log ED50 (%RSE)
Weighting by N	-0.018 (65.6%)	-4.587 (21.2%)	3.404 (39.8%)
Replace all* SDs by weighted mean of available SDs	-0.018 (46.6%)	-4.587 (15.1%)	3.404 (28.3%)
Replace all* SDs by maximum of available SDs	-0.018 (64.7%)	-4.587 (20.9%)	3.404 (39.3%)
Impute missing SDs by weighted mean of available SDs	-0.016 (49.7%)	-4.331 (15.6%)	3.290 (32.1%)
Impute missing SDs by maximum of available SDs	-0.012 (74.0%)	-4.864 (21.9%)	3.862 (27.1%)
Impute missing SD by random forest	-0.016 (46.8%)	-4.330 (16.3%)	3.345 (32.1%)

Table 4. Assessment of Methods on Observed Data

Weighting by N has generally inflated %RSEs

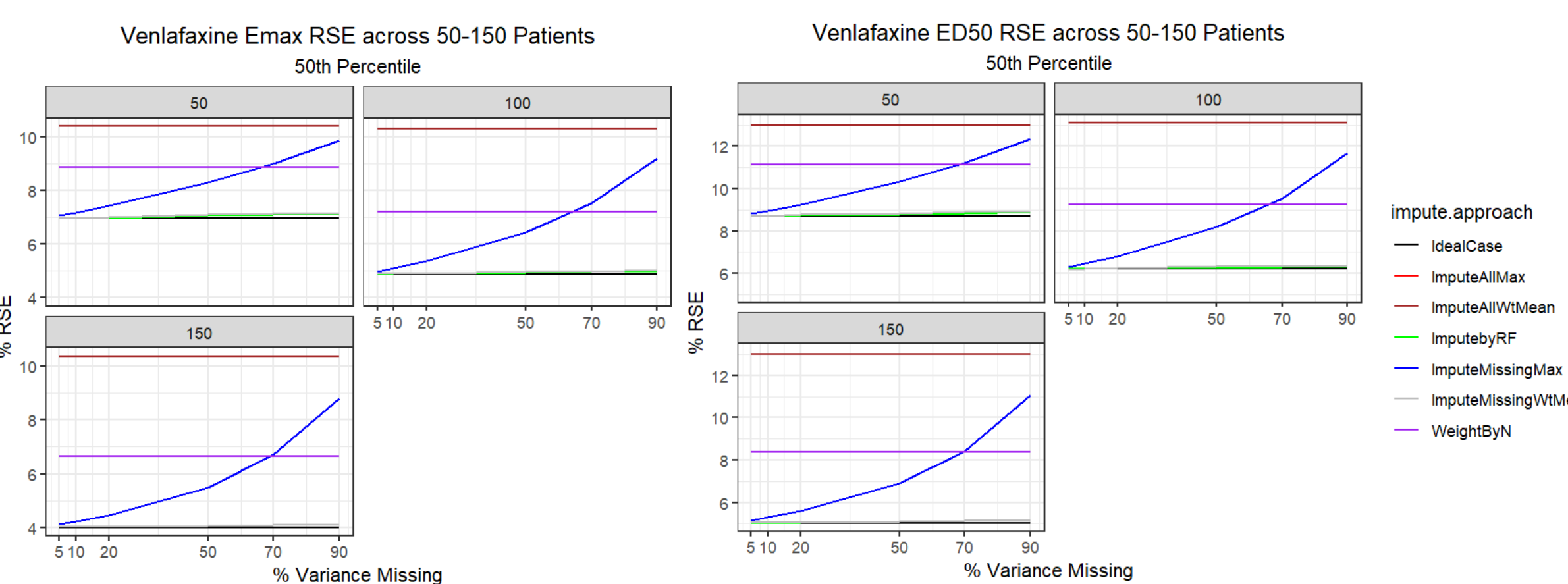


Figure 4. Median precision for Emax, ED50 & slope - across levels of missingness, faceted by sample size, colored by approach + ideal case

(Ideal case is when all resampled SDs were used as fixed weights in simulations)

Imputing missing variance by RF and weighted mean have the lowest %RSE across levels of missingness, while replacing all by weighted mean and maximum have the highest %RSE.

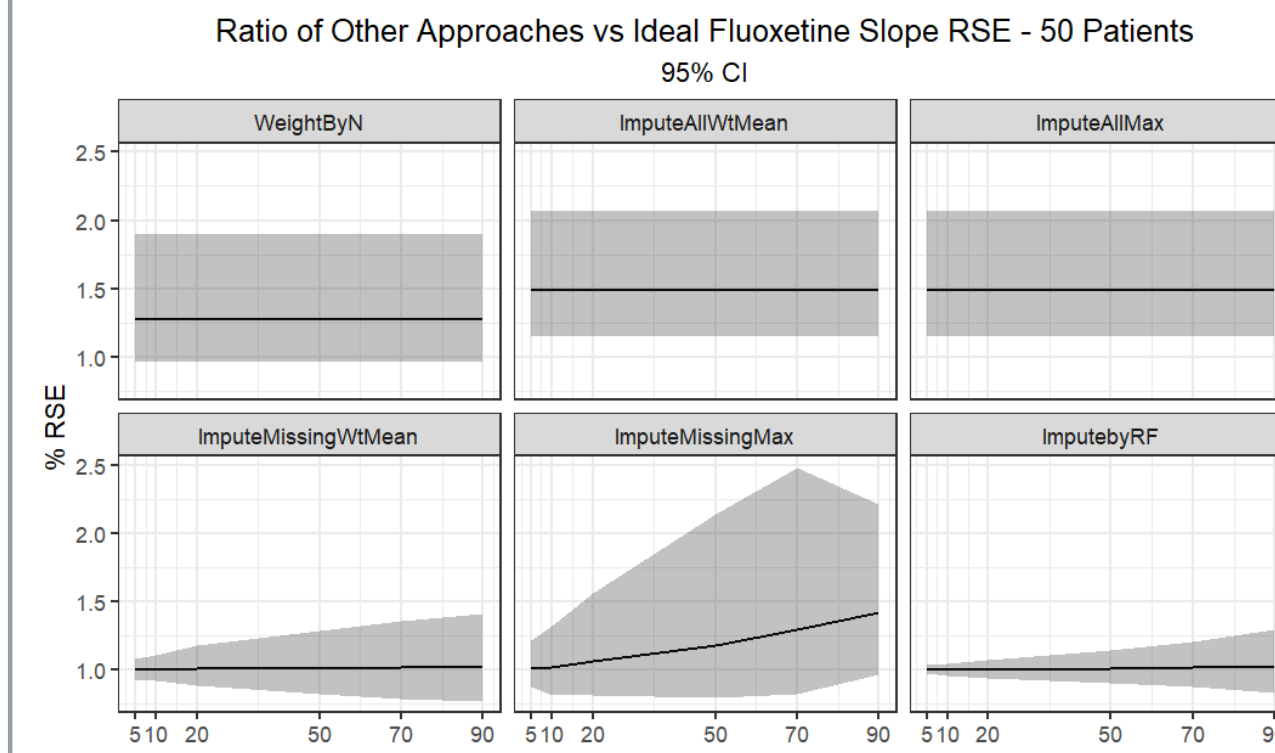


Figure 5. Median precision ratio of methods against ideal case (with 95% CI) - across levels of missingness, faceted by approach (N = 50 only)

Imputing missing variance by RF and weighted mean have median ratios closest to 1 across all % missingness. Imputing by RF has an even narrower 95% CI, indicating its highest precision.

SUMMARY

Observed data results

- Parameter estimates seem unimpacted by method, except for those based on replacing missing SDs by maximum, which appear to differ slightly.
- Weighting by N appears to have inflated %RSEs, comparable to the worst-case scenario wherein missing SDs were replaced by the maximum available SD.

Simulated data results

- Imputing missing by weighted mean and by RF tend to have the lowest median %RSE across levels of missingness - in line with the ideal case.
- At around 70% missingness, the median %RSE of imputing missing by maximum became comparable to that of weighting by N, further implying that the latter approach may be conservative.
- Median %RSE ratios from imputing missing by weighted mean and by RF - both against the ideal case - are the closest to 1 (lowest).
- The 95% CI of imputing by RF is the smallest across %missingness, indicating highest precision.

Bias, variance, and mean squared error (MSE) of estimates were also examined, but no conclusive patterns were observed on impact on parameters across the methods.

When some measures of precision are reported, it is recommended to impute missing values using the weighted mean of available information or by RF as use as fixed weights.

LIMITATIONS & FUTURE WORK

- Further investigation into imputation by RF is necessary because the RF in these simulations may not fully capture correlations in the original data.
- Since 49% of the original precisions were missing, parameter estimates from the RF approach were selected as "true" values and used as the basis for simulations. An optimal scenario is when all measures of precision are available.
- Further investigation into the impact of approach or level of missingness on bias, variance and MSE is required.

REFERENCES

- Kambach S, Bruehlheide H, Gerstner K, Gurevitch J, Beckmann M, Seppelt R. Consequences of multiple imputation of missing standard deviations and sample sizes in meta-analysis. *Ecology and Evolution*. 2020;10(20):11699-11712. doi:https://doi.org/10.1002/ece3.6806
- Chowdhry AK, Dworkin RH, McDermott MP. Meta-analysis with missing study-level sample variance data. *Statistics in Medicine*. 2016;35(17):3021-3032. doi:https://doi.org/10.1002/sim.6908