



Comparison of Methods for Handling Missing Covariate Data

Åsa M. Johansson, Mats O. Karlsson

Dept of Pharmaceutical Biosciences, Uppsala University, Sweden

Background

Missing covariate data is a common problem in non-linear mixed effects modelling and the method chosen for handling missing data can be crucial for the outcome of the study. Missing data can be divided into three categories: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR) [1]. The underlying mechanism of the missingness is usually unknown but will affect the predictability of the model if wrong assumptions are made.

Objective

To implement and compare different methods for handling missing categorical covariate data under different mechanisms of missingness.

Methods

Simulation model:

A constant infusion PK model with a 100% difference in CL between males and females, an inter-individual variability of 30% and a residual error of 20%.

- Simulations generated data for 200 individuals (60% males)
- Weights were simulated according to sex specific truncated log-normal distributions
- 50% of the individuals were missing information about the covariate sex
 - MCAR – Equal probability of missing sex for all individuals
 - MAR – Higher probability of missing sex with increasing weight
 - MNAR – Three times higher probability of missing sex for males

Handling of missing data:

Multiple Imputation (MI)

A modified version of the method described by Wu and Wu [2].

- 1) A NONMEM model without covariates was created
- 2) Individual (empirical Bayes) estimates of CL (CL_i) were obtained
- 3) A model was created to describe the probability of being male given the observed weight (WT) and the estimated CL_i ($p(\text{male}|\text{WT}, CL_i)$)
- 4) m ($m=6$) sets of complete datasets were created through repeated sampling of sex from the probability model in step 3
- 5) Final estimates of CL were obtained by taking the average of the estimates in step 4

Modelling with \$MIX based on observed weight (MOD)

- 1) A model was created to describe the probability of being male given the observed weight ($p(\text{male}|\text{WT})$)
- 2) The most probable value of sex was modelled in \$MIX using the probability model in step 1

Modelling with \$MIX based on observed weight with estimation of additional fixed effect (EST)

- 1) A model was created to describe the probability of being male given the observed weight ($p(\text{male}|\text{WT})$)
- 2) The most probable value of sex was modelled in \$MIX using the probability model in step 1 with estimation of additional fixed effect

Stochastic Simulations and Estimations (SSE): The simulation model was used to simulate 200 datasets. The datasets were then analyzed and the missingness were handled according to MI, MOD and EST. For comparison purposes, estimation with all data (ALL) was also carried out.

Bias and precision: The difference (Δ) between final θ estimate and true parameter value was calculated for each parameter and each dataset. Bias and precision were evaluated by calculation of RRMSE, scaled by the fraction of males and females respectively, and in box plots showing the spread and the deviation from zero of Δ for each parameter and each method under each mechanism of missingness.

Conclusions

- ❖ MI and/or MOD/EST may all be appropriate methods to handle missing covariates but they differ in their robustness to misspecification of the relation to known covariates, missingness mechanism and data richness.
- ❖ A significant drop in OFV between MOD and EST is an indication of data being MNAR and/or a probability equation with poor predictability.
- ❖ Bias and imprecision in the implemented sex to weight relationship may have a substantial effect on the PK parameter estimates for all underlying mechanisms of missingness.
- ❖ The MI method used shows unbiased and precise results for MCAR and MAR but the method assumes low EBE shrinkage.

Results

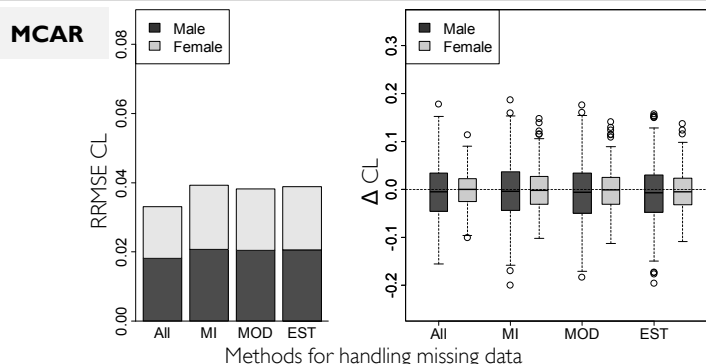


Figure 1: The bar plot shows RRMSE scaled by the fraction of males and females respectively, and the box plot shows the bias and precision of the fixed effect in CL for males and females respectively, when the underlying mechanism of missingness is MCAR.

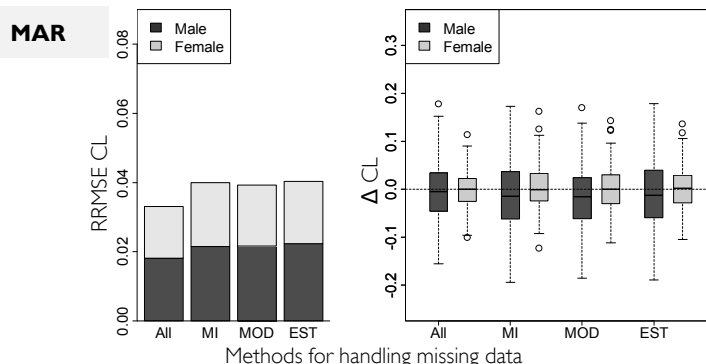


Figure 2: The bar plot shows RRMSE scaled by the fraction of males and females respectively, and the box plot shows the bias and precision of the fixed effect in CL for males and females respectively, when the underlying mechanism of missingness is MAR.

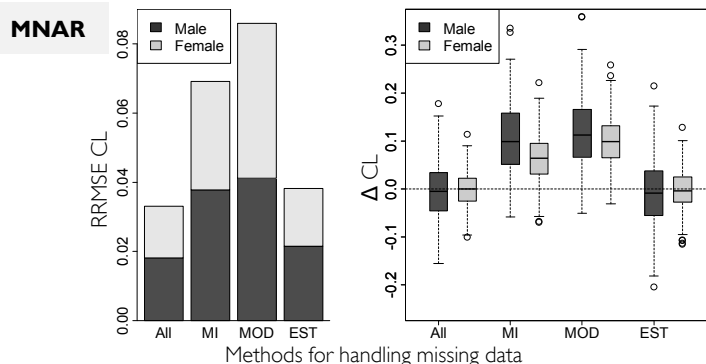


Figure 3: The bar plot shows RRMSE scaled by the fraction of males and females respectively, and the box plot shows the bias and precision of the fixed effect in CL for males and females respectively, when the underlying mechanism of missingness is MNAR.

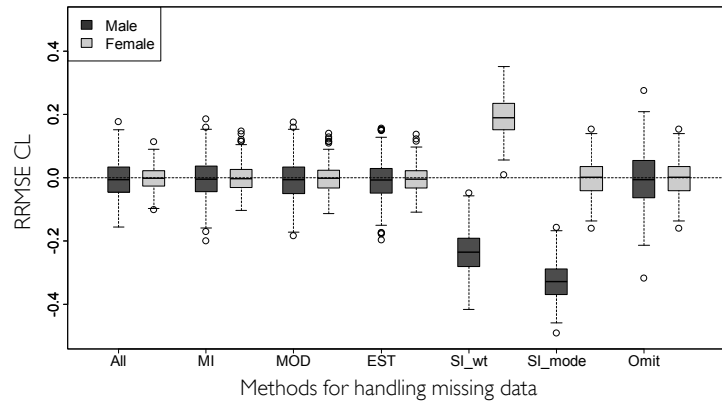
[1] Little and Rubin. Statistical analysis with missing data, 2002

[2] Wu and Wu. Statistics in Medicine, 2001

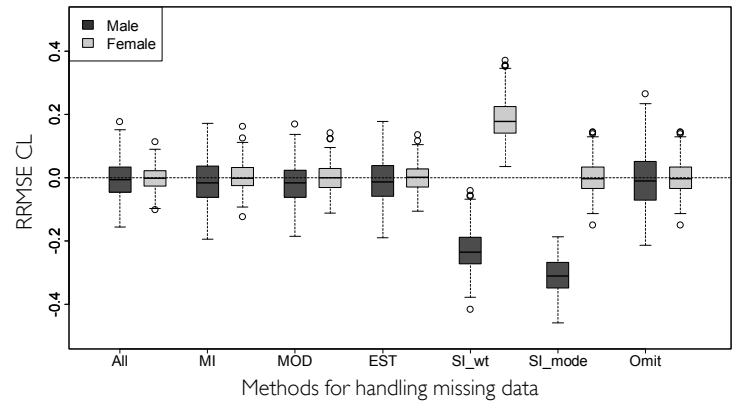
Methods for handling missing data

- MI – Multiple imputation based on weight and EBEs
- MOD – Modelling with \$MIX based on weight (logit-transformed linear regression)
- EST – Modelling with \$MIX based on weight with estimation of additional fixed effect
- SI_wt – Single imputation based on weight
- SI_mode – Single imputation of mode
- Omit – Complete case analysis

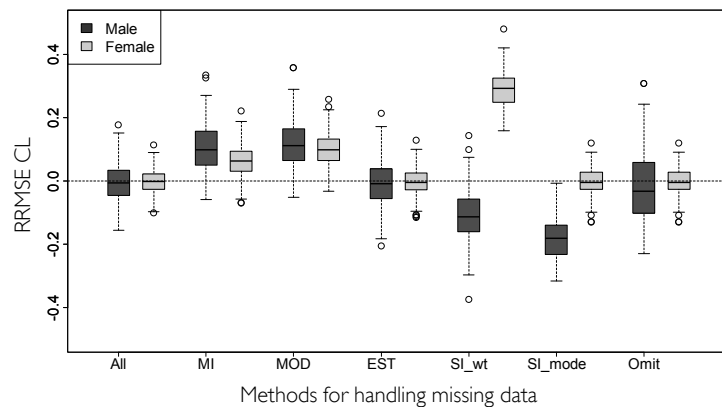
MCAR



MAR



MNAR



Probability density function based on log-normal weight distributions

NONMEM code (imputation):

\$PRED

IF(ICALL == 4) THEN

IMALE = 0

IF(NEWIND /= 2) CALL RANDOM (2, R)

PM1 = PHI(LOG(WT/85.1)/SQRT(0.0329))

PM2 = PHI(LOG((WT-0.1)/85.1)/SQRT(0.0329))

PF1 = PHI(LOG(WT/73.0)/SQRT(0.0442))

PF2 = PHI(LOG((WT-0.1)/73.0)/SQRT(0.0442))

PM = 0.6*(PM1 - PM2)

PF = 0.4*(PF1 - PF2)

PMALE = PM/(PM + PF)

IF(R <= PMALE) IMALE = 1

ENDIF

- ❖ Gives precise and unbiased results as long as the estimations of mean and variance of the weight distributions are precise and unbiased.
- ❖ PM1 is the value of the cumulative distribution function for males with the weight WT, given the mean weight 85.1 kg and the variance 0.0329 kg².
- ❖ PM2 is the value of the cumulative distribution function for males with the weight WT-0.1.
- ❖ PM is the value of the probability density function (PM1-PM2) for males with the weight WT, scaled with the fraction of males in the data.
- ❖ PF1, PF2 and PF are the corresponding values of the cumulative distribution functions and the probability density function for females (mean weight 73.0 kg and variance 0.0442 kg²).
- ❖ PMALE is the probability of being male given the weight WT.