EP16: Missing Values in Clinical Research: Multiple Imputation

11. Imputation with Longitudinal Data

Nicole Erler

Department of Biostatistics, Erasmus Medical Center

n.erler@erasmusmc.nl

Erasmus MC
University Medical Center Rotterdam

The Challenge with Longitudinal Data

In **long format**:

- (potential) correlation between repeated measurements ignored
- inconsistent imputation of baseline covariates

The Challenge with Longitudinal Data

In **long format**:

- (potential) correlation between repeated measurements ignored
- inconsistent imputation of baseline covariates

In **wide format**:

- for **unbalanced** data: direct wide format may not be possible
- for relatively **balanced** data: inefficient

The Challenge with Longitudinal Data

In **long format**:

- (potential) correlation between repeated measurements ignored
- inconsistent imputation of baseline covariates

In **wide format**:

- for **unbalanced** data: direct wide format may not be possible
- for relatively **balanced** data: inefficient

Simple summaries to allow wide format:

- loss of information
- potential MNAR
- **bias**

Imputation with mice

mice has functions to allow imputation of longitudinal (2-level) data:

- **Level 1:**
  repeated measurements within subjects or subjects within classes
- **Level 2:**
  time-constant/baseline covariates, between subjects effects, variables on the group level

Imputation methods for **level-1** variables:

- `2l.pan`
- `2l.norm`
- `2l.lmer`
- `2l.bin`

Imputation methods for **level-2** variables:

- `2lonly.norm`
- `2lonly.pmm`
- `2lonly.mean`

Imputation with mice

- `2l.pan`: linear two-level model with **homogeneous within group variances** (Schafer and Yucel 2002)
- `2l.norm`: (Bayesian) linear two-level model with **heterogenous** group variances

Imputation with mice

- `2l.pan`: linear two-level model with **homogeneous within group variances** (Schafer and Yucel 2002)
- `2l.norm`: (Bayesian) linear two-level model with **heterogenous** group variances

- `2l.lmer`/`2l.bin`: univariate systematically and sporadically missing data using a two-level normal/logistic model using `lmer()`/`glmer()` from package **lme4**.

Imputation with mice

- `2l.pan`: linear two-level model with **homogeneous within group variances** (Schafer and Yucel 2002)
- `2l.norm`: (Bayesian) linear two-level model with **heterogenous** group variances

- `2l.lmer`/`2l.bin`: univariate systematically and sporadically missing data using a two-level normal/logistic model using `lmer()`/`glmer()` from package **lme4**.

- `2lonly.norm` and `2lonly.pmm`: to impute level-2 variables

Imputation with mice

- `2l.pan`: linear two-level model with **homogeneous within group variances** (Schafer and Yucel 2002)
- `2l.norm`: (Bayesian) linear two-level model with **heterogenous** group variances

- `2l.lmer`/`2l.bin`: univariate systematically and sporadically missing data using a two-level normal/logistic model using `lmer()`/`glmer()` from package **lme4**.

- `2lonly.norm` and `2lonly.pmm`: to impute level-2 variables

- `2lonly.mean`: imputes values with the mean of the observed values per class (only to be used in special cases!)

Imputation with mice

The `predictorMatrix` contains extra info for multi-level imputation:

- grouping/ID variable: -2
- random effects (also included as fixed effects): 2
- fixed effects of group means: 3
- fixed effects of group means & random effects: 4

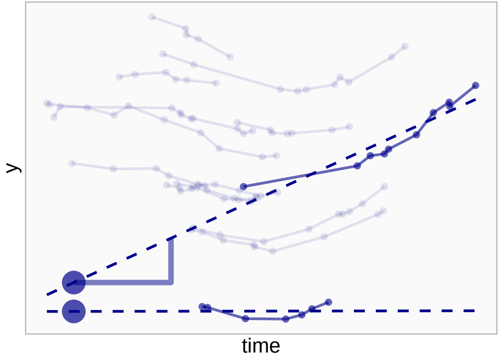
In all cases, the group identifier ("id" variable) needs to be set to -2 in the `predictorMatrix`.

Imputation with mice

Alternative approach: (Erler et al. 2016)
Get a **better summary** of the longitudinal variables!

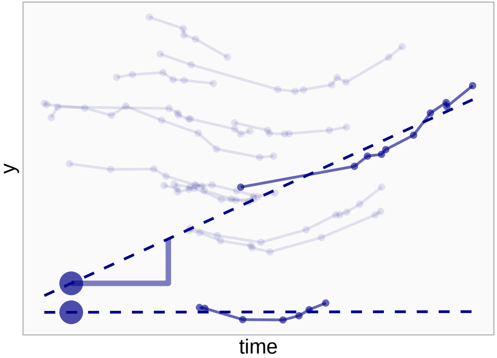
Imputation with mice

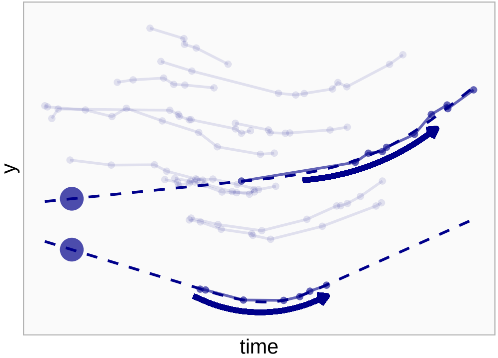
Alternative approach: (Erler et al. 2016)
Get a **better summary** of the longitudinal variables!



Imputation with mice

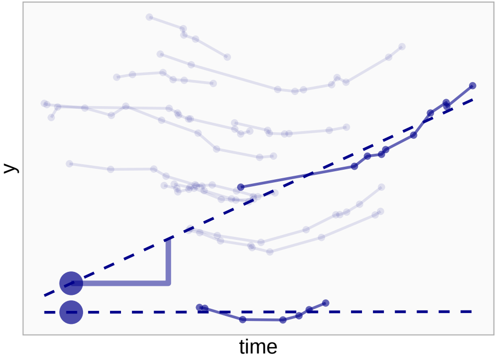
Alternative approach: (Erler et al. 2016)
Get a **better summary** of the longitudinal variables!

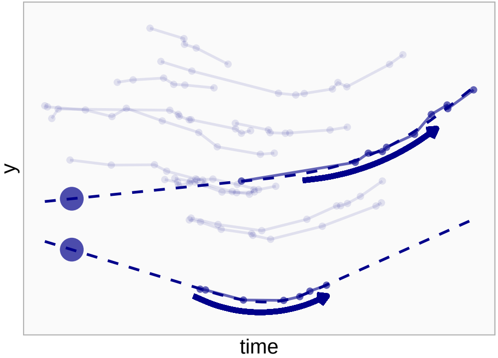




Imputation with mice

Alternative approach: (Erler et al. 2016)
Get a **better summary** of the longitudinal variables!





Approximate trajectories using random effects!

Imputation with mice

- Fit a flexible mixed model for outcome and other longitudinal variables
  ➡ extract the random effects

Imputation with mice

- Fit a flexible mixed model for outcome and other longitudinal variables
  ➡ extract the random effects

- reduce data to a “baseline” version & add random effects as columns

Imputation with mice

- Fit a flexible mixed model for outcome and other longitudinal variables
  ➡ extract the random effects
- reduce data to a “baseline” version & add random effects as columns
- impute with **mice**

Imputation with mice

- Fit a flexible mixed model for outcome and other longitudinal variables
  ➡ extract the random effects
- reduce data to a “baseline” version & add random effects as columns
- impute with **mice**
- extract imputed data & merge with long. variables

Imputation with mice

- Fit a flexible mixed model for outcome and other longitudinal variables
  ➡ extract the random effects
- reduce data to a “baseline” version & add random effects as columns
- impute with **mice**
- extract imputed data & merge with long. variables
- convert back to `mids` object & analyse

Imputation with mice

- Fit a flexible mixed model for outcome and other longitudinal variables
  ➡ extract the random effects
- reduce data to a “baseline” version & add random effects as columns
- impute with **mice**
- extract imputed data & merge with long. variables
- convert back to `mids` object & analyse

Imputation with mice

- Fit a flexible mixed model for outcome and other longitudinal variables
  ➡ extract the random effects
- reduce data to a "baseline" version & add random effects as columns
- impute with **mice**
- extract imputed data & merge with long. variables
- convert back to `mids` object & analyse

Drawback: cannot handle **incomplete longitudinal variables**.

Imputation with JointAI

Example data:

- *x*1 (complete)
- *x*2 (binary, 30% NA)
- *x*3 (3 categories, 30% NA)
- *x*4 (continuous/normal, 30% NA)
- *y* (longitudinal outcome)
- *time* (time variable with quadratic effect)
- *id* (id variable)

Imputation with JointAI

The syntax for analysing mixed models in **JointAI** is analogous the syntax used in `lme()` of the package **nlme**.

```
library("JointAI")
JointAI_long <- lme_imp(y ~ x1 + x2 + x3 + x4 + time + I(time^2),
                        random = ~time|id, data = longDF2,
                        n.iter = 5000)
```

Imputation with JointAI

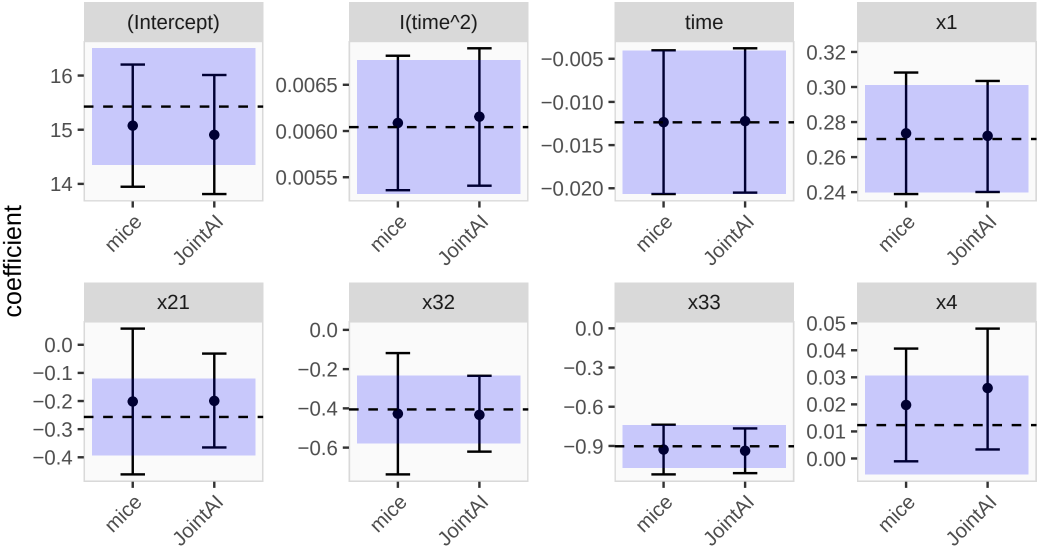
The syntax for analysing mixed models in **JointAI** is analogous the syntax used in `lme()` of the package **nlme**.

```
library("JointAI")
JointAI_long <- lme_imp(y ~ x1 + x2 + x3 + x4 + time + I(time^2),
                        random = ~time|id, data = longDF2,
                        n.iter = 5000)
```

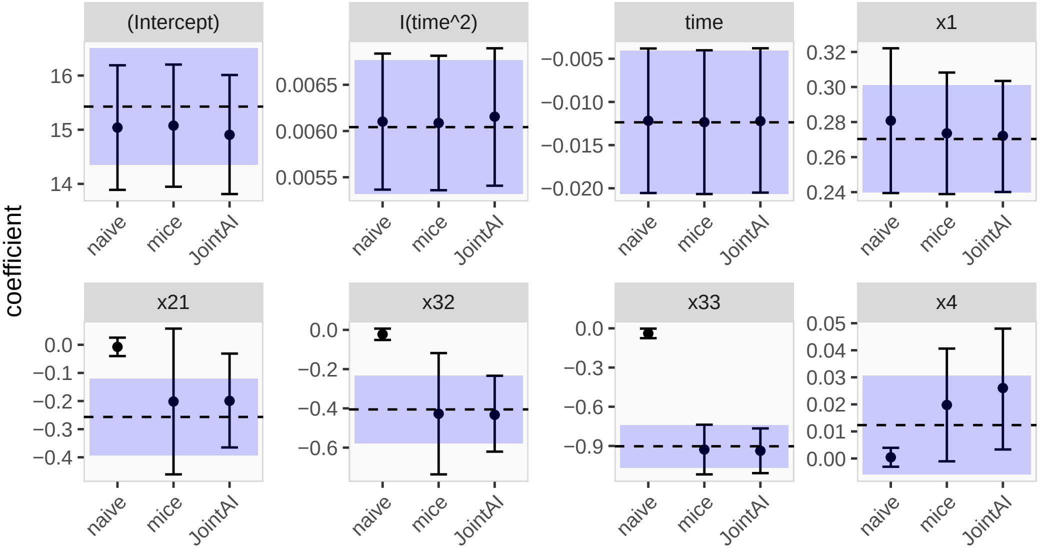
Again, convergence of the Gibbs sampler needs to be checked before obtaining the results.

Contrary to the two-level imputation of **mice**, non-linear associations are appropriately handled.

Comparison of Results



Comparison of Results



References

Erler, Nicole S, Dimitris Rizopoulos, Joost van Rosmalen, Vincent WV Jaddoe, Oscar H Franco, and Emmanuel MEH Lesaffre. 2016. "Dealing with Missing Covariates in Epidemiologic Studies: A Comparison Between Multiple Imputation and a Full Bayesian Approach." *Statistics in Medicine* 35 (17): 2955–74. https://doi.org/10.1002/sim.6944.

Schafer, Joseph L, and Recai M Yucel. 2002. "Computational Strategies for Multivariate Linear Mixed-Effects Models with Missing Values." *Journal of Computational and Graphical Statistics* 11 (2): 437–57.