An Alternative Characterization of Missing At Random in Shared Parameter Models

G. Papageorgiou[1], D. Rizopoulos[1]

1. Department of Biostatistics, Erasmus MC, Rotterdam, The Netherlands

August 25th, 2020

ISCB 2020

- Dropout is a common complication in longitudinal studies

- Missing Data Mechanisms:

  - **MCAR**: Missingness does not depend on neither observed nor unobserved outcomes

  - **MAR**: Missingness is independed of the unobserved outcomes after conditioning on observed outcomes

  - **MNAR**: Missingness depends on unobserved outcomes

- **MCAR and MAR are ignorable**:

  - Under the (Bayesian) Likelihood Framework

  - If missingness and measurement processes depend on different sets of parameters

- Distinction between MAR and MNAR is intractable (Molenberghs et al., 2007)

- Start with MAR valid analysis $\rightarrow$ Explore MNAR deviations

- Three frameworks for sensitivity analysis:

  - **Selection Models**: Naturally encompass MAR

  - **Pattern Mixture Models**: MAR characterization for longitudinal and time-to-event data (Molenberghs et al. 1998)

  - **Shared Parameter Models**: Generalized Shared Parameter Model (Creemers et al. 2010, 2011; Njagi et al. 2014)

- Focus on Shared Parameter Models:

  - MNAR and MAR on the **subject specific-level**

  - **Sensitivity analysis** using joint models for longitudinal data and dropout

- The shared parameter model is defined as:

$$p\left(y_i^o, y_i^m, T_i^*, C_i; \theta, \psi\right) = \int p\left(y_i^o, y_i^m, T_i^*, C_i, b_i; \theta, \psi, D\right) db_i,$$

- $y_i = \left(y_{i1}, \dots, y_{in_i}\right)^{\top}$

- $y_i = \left(y_i^o, y_i^m\right)$

- $T_i = \min\left(T_i^*, C_i\right),$

- δ_i: dropout indicator.

- $b_i \sim \mathcal{N}\left(0, D\right)$

- Assumptions:

  - Conditional Independence

  - Non Informative Censoring

$$\int p\left(T_i^* \mid b_i; \psi^{T^*}\right) p\left(C_i \mid y_i^o; \psi^C\right) p\left(y_i^o, y_i^m \mid b_i; \theta\right) p\left(b_i; D\right) db_i,$$

- On the subject specific level:

$$
\begin{cases}
\displaystyle\int p\left(T_i \mid b_i; \psi^{T^*}\right) p\left(y_i^o, y_i^m \mid b_i; \theta\right) p\left(b_i; D\right) db_i, & i : \text{dropout} \quad \rightarrow \quad \textbf{MNAR} \\[2em]
\displaystyle\int p\left(C_i \mid y_i^o; \psi^C\right) p\left(y_i^o, y_i^m \mid b_i; \theta\right) p\left(b_i; D\right) db_i, & i : \text{censored} \quad \rightarrow \quad \textbf{MAR}.
\end{cases}
$$

- Joint model for longitudinal data and dropout:

$$
\begin{cases}
\begin{aligned}
y_i\left(t\right) &= \eta_i\left(t\right) + \epsilon_i\left(t\right) \\
&= x_i^\top\left(t\right)\beta + z_i^\top b_i + \epsilon_i\left(t\right) \\[1em]
h_i\left(t\right) &= h_0\left(t\right)\exp\left\{w_i^\top\gamma_i + \alpha\eta_i\left(t\right)\right\}
\end{aligned} \\[2em]
\qquad b_i \sim \mathcal{N}\left(0, D\right), \epsilon_i \sim \mathcal{N}\left(0, \sigma^2\right)
\end{cases}
$$

- We can perturb the dropout indicator δ_i to explore different scenarios

- HIV CD4 Data (Abrams et al. 1994; Goldman et al. 1996):

  - Randomized Clinical Trial

  - 467 HIV patients randomised to Didanosine (ddI) and Zalcitabine (ddC)

  - Scheduled visits at baseline, 2, 6, 12 and 18 months

  - 184 (39%) subjects died

 - around 60% of the planned measurements were actually recorded
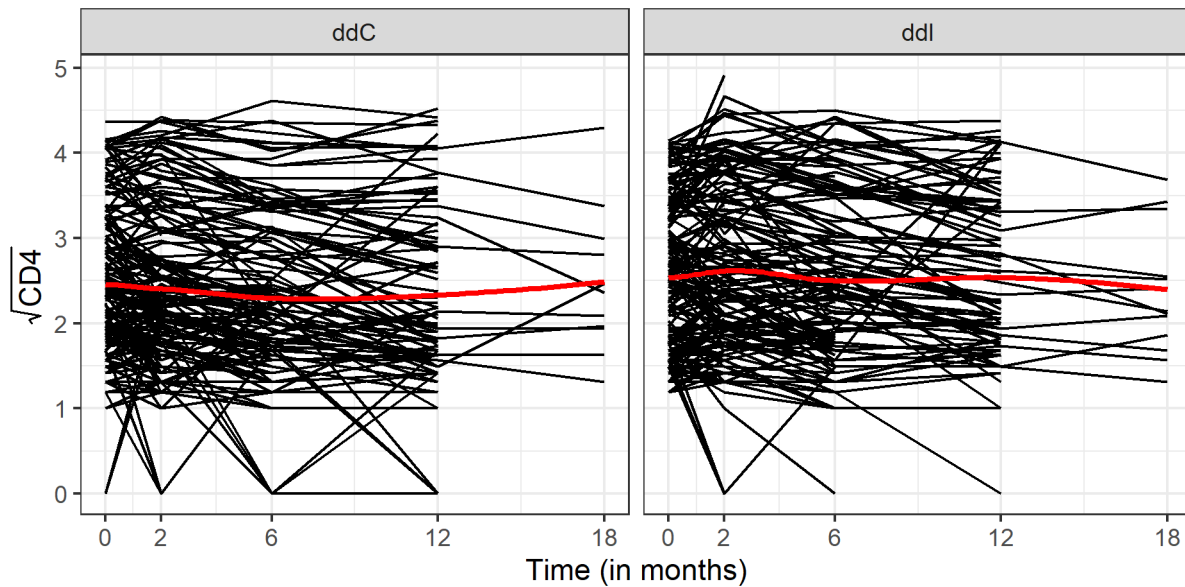
| Dropout Pattern	ddC		ddI	
	N	%	N	%
OXXXX	29	14.4%	32	15.6%
OOXXX	35	17.4%	37	18.0%
OOOXX	41	20.4%	47	22.9%
OOOOX	85	42.3%	76	37.1%
OOOOO	11	5.5%	13	6.3%
Total	201	100%	205	100%

- Joint model for longitudinal and dropout processes:

$$\begin{cases} \sqrt{y_{ij}^o} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1}) t_{ij} + \beta_2 \mathsf{ddI} + \beta_3 \left(t_{ij} \times \mathsf{ddI} \right) + \epsilon_{ij}, \\[2mm] h_i \left(t \mid \mathcal{H}_i \left(t \right), b_i \right) = h_0 \left(t \right) \exp \left[\gamma_1 \mathsf{ddI} + \alpha \eta_i \left(t \right) \right] \\[2mm] b_i \sim \mathcal{N} \left(0, D \right), \epsilon_i \sim \mathcal{N} \left(0, \sigma^2 \right) \end{cases}$$

- Three models:

  - **MNAR(1)**: No distinction between death and other causes

  - **MNAR(2)**: Distinction between death and other causes

  - **MAR**: All cases as censored

Results

| Effect	Parameter	MNAR(1)	MNAR(2)	MAR
Intercept	β_0	2.4671 (0.0640)	2.4680 (0.0629)	2.4423 (0.0647)
Time	β_1	-0.0655 (0.1701)	-0.0674 (0.1715)	-0.0399 (0.0050)
ddI	β_2	0.1038 (0.0908)	0.1067 (0.0879)	0.1188 (0.0946)
ddI $\times$ Time	β_3	0.0219 (0.2332)	0.0290 (0.2376)	0.0089 (0.0070)
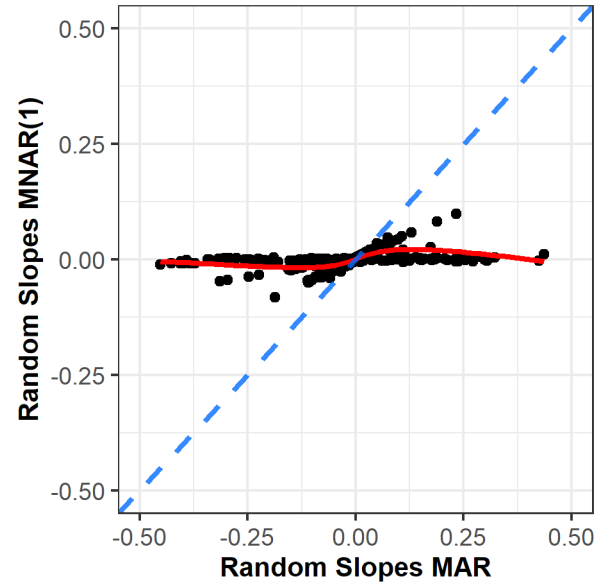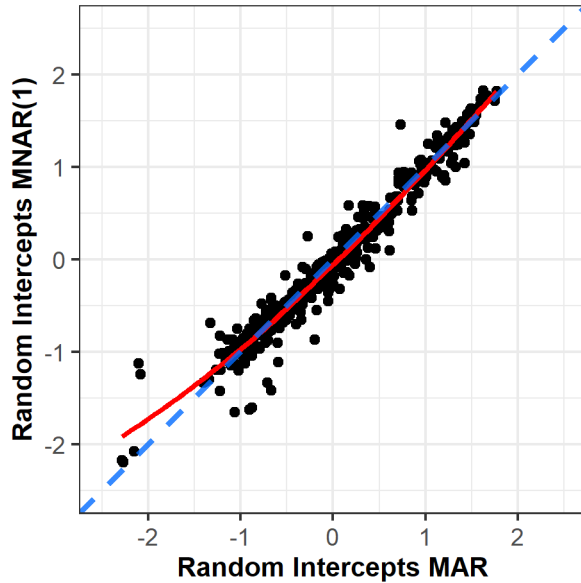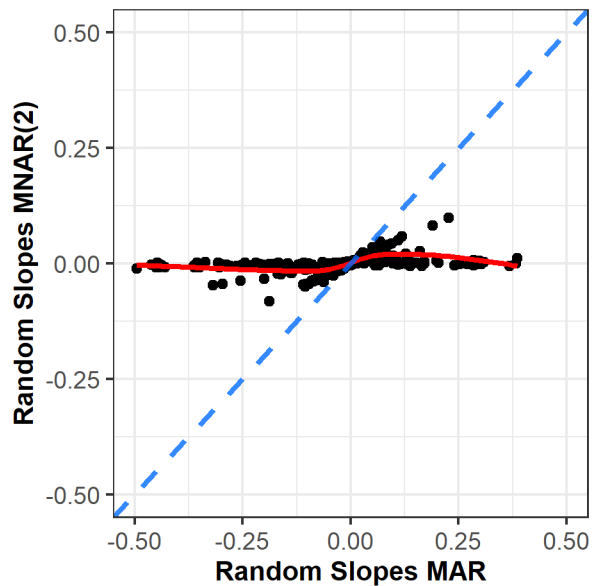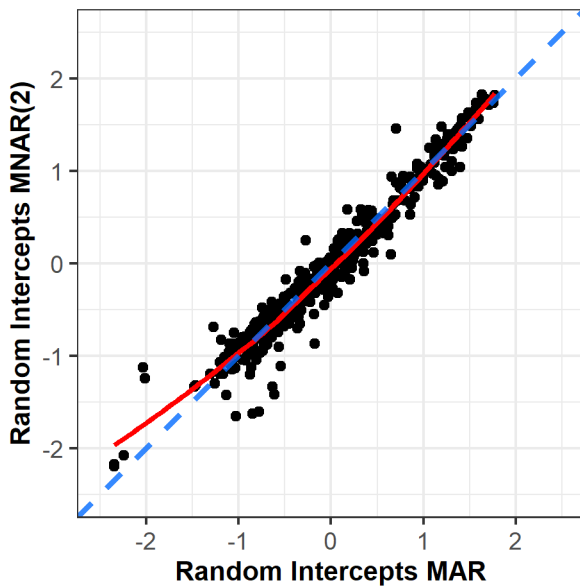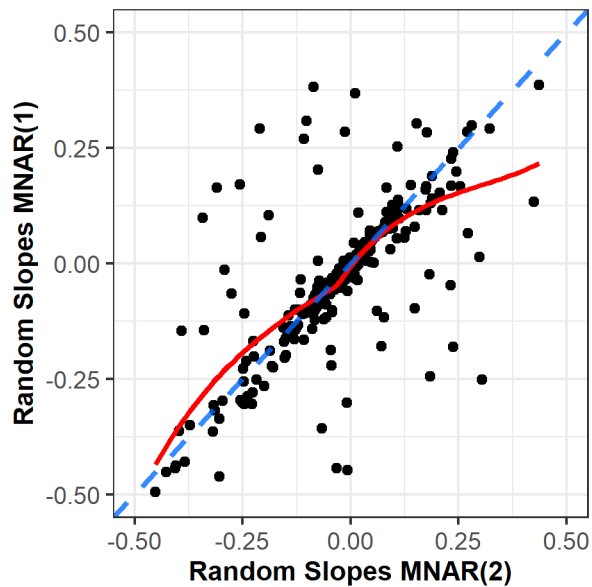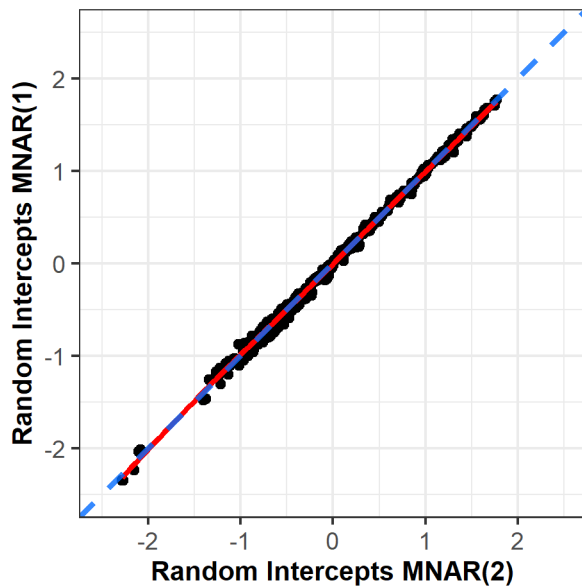
- Scenarios:

  - 5% MNAR - 45% MAR

  - 25% MNAR - 25% MAR

  - 45% MNAR - 5% MAR

- Models:

  - **MAR**: all dropout cases are considered as MAR

  - **MNAR(1)**: all dropout cases are considered MNAR

  - **MNAR(2)**: dropout cases are considered MNAR or MAR depending on the reason of dropout
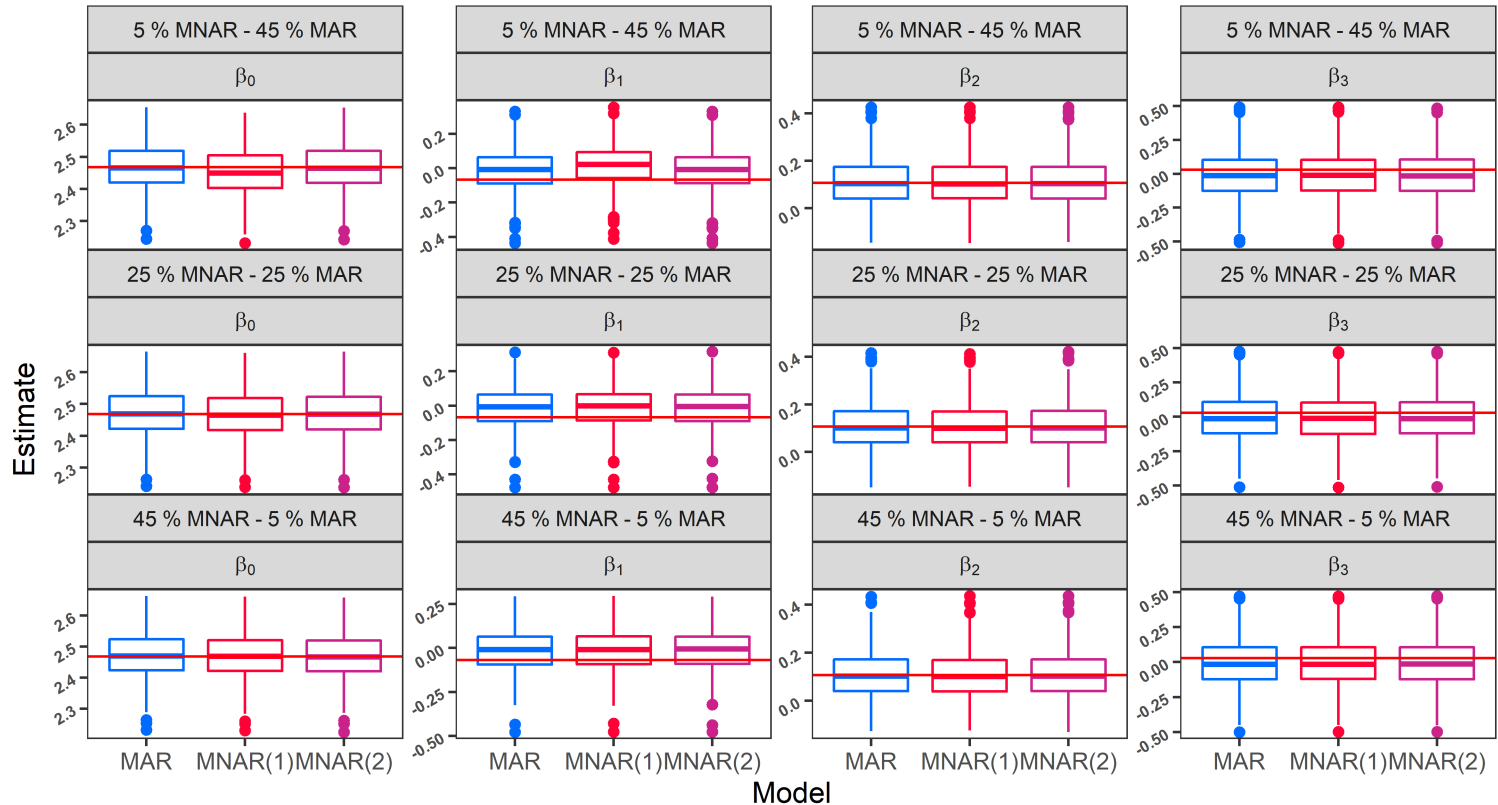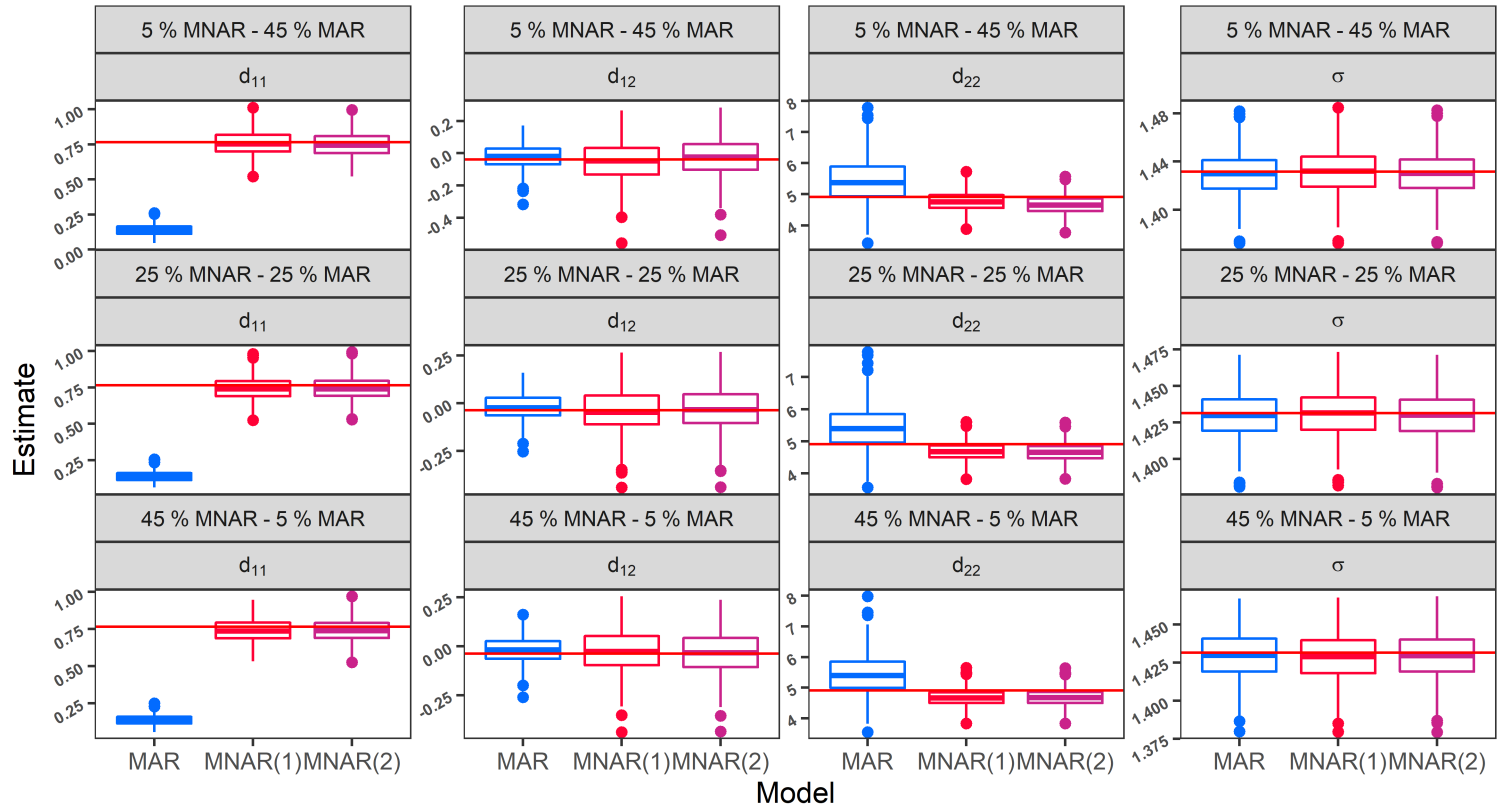
- Sensitivity analysis should be routinely reported for longitudinal studies with dropout

  - SPMs can be used to explore MAR and MNAR scenarios

  - The main advantage is that this can be done at a subject-specific level

  - This can be achieved with existing software such as **JMbayes**

  - Different causes of missingness can be explored if information is available.

- Full paper available online:
  - Papageorgiou, G., & Rizopoulos, D. (2020). An alternative characterization of MAR in shared parameter models for incomplete longitudinal data and its utilization for sensitivity analysis. Statistical Modelling. **https://doi.org/10.1177/1471082X20927114**

Thank you

@gr_papageorgiou

https://github.com/gpapageorgiou

g.papageorgiou@erasmusmc.nl

gregpapageorgiou.com