11. Missing values

Sonja Greven

Summer Term 2017

Overview Chapter 11 - Missing values

11.1 Introduction

11.2 Three different missing value mechanisms

11.3 Special case: dropout

11.4 ML based methods and GEE for missing values

11.5 Overview of data analysis methods for missing values

11.6 Models for the dropout process

Missing data

- Missing data is common in longitudinal studies. Data is missing if a measurement that was intended to be taken is not taken, or not available for another reason.

- The reason for missing measurements is important. For example:
  - The lab technician accidentally destroyed the blood sample.
  - Measurements below the limit of detection are set to missing (censoring).
  - The values are missing because the subjects did not show up for their scheduled visits.

Notation

Assumption: It is planned to take $n_i = n$ measurements per subject.

- Vector of responses (observed and missing) for subject i:

$$\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in})^T$$

- $R_{ij} = 1$, if Y_{ij} is observed, otherwise $R_{ij} = 0$. For each subject a vector

$$\mathbf{R}_i = (R_{i1}, \ldots, R_{in})^T$$

  is obtained.

- $\mathbf{R}_i$ results in a division of $\mathbf{Y}_i$ into two components $\mathbf{Y}_i^o$ (observed) and $\mathbf{Y}_i^m$ (missing).
- Subjects with $R_{ij} = 1$ for all j (i.e. without missing values) are called completers.

Missing data patterns

- Dropout / loss-to-follow-up / attrition:

  Whenever Y_{ij} is missing, so are all Y_{ik} for $k \geq j$.
  Pattern: $\boldsymbol{R}_i = (R_{i1}, \ldots, R_{i(D_i-1)}, R_{iD_i}, \ldots, R_{in})^T = (1, \ldots, 1, 0, \ldots, 0)^T$
  with dropout indicator

$$D_i = 1 + \sum_{j=1}^{n} R_{ij}.$$

- Intermittent missing values

  Example patterns: $\boldsymbol{R}_i = (1, 1, 0, 1, \ldots, 1)^T$, $\boldsymbol{R}_i = (1, 0, 1, 0, 1, \ldots)^T$.

Questions

For missing values, is it allowed to:

- calculate means and variances?
- use ML based methods?
- use the GEE method?

Important: The answer for each method depends on the missing mechanism

- missing completely at random (MCAR)
- missing at random (MAR)
- not missing at random (NMAR)

(Rubin, 1976)

Overview Chapter 11 - Missing values

11.1 Introduction

11.2 Three different missing value mechanisms

11.3 Special case: dropout

11.4 ML based methods and GEE for missing values

11.5 Overview of data analysis methods for missing values

11.6 Models for the dropout process

Missing completely at random (MCAR)

$$P(R_{ij} = 1|\mathbf{Y}_i^o, \mathbf{Y}_i^m, \mathbf{X}_i) = P(R_{ij} = 1|\mathbf{X}_i).$$

for $i = 1, \ldots, N$, $j = 1, \ldots, n$.

- The probability of missingness ($P(R_{ij} = 0)$) is not related to any of the responses. The distribution of the Y_{ij} is the same as that of the Y_{ij}^o, given $\boldsymbol{X}_i$.

- Other (stronger) definition: also no connection between the covariates and the occurrence of missing values,

  $$P(R_{ij} = 1|\mathbf{Y}_i^o, \mathbf{Y}_i^m, \mathbf{X}_i) = P(R_{ij} = 1).$$

  The observed data are a random sample of the complete data.

Examples MCAR

- The lab technician accidentally destroyed the blood sample.
- Overlooked question on questionnaire
- Questionnaire lost in the mail
- Did not come to examination because of a death in the family
- Rotating panel: patients by design rotate out of the study after providing a pre-determined number of measurements.
- Death due to a car accident
- Moving, but with exceptions

$\rightarrow$ Try to find out from data collector

Examples MCAR

Example for

$$P(R_{ij} = 1|\mathbf{Y}_i^o, \mathbf{Y}_i^m, \mathbf{X}_i) = P(R_{ij} = 1|\mathbf{X}_i).$$

Weight and sex: suppose that regardless of the weight itself women hesitate to give their weight:

$$P(R_{ij} = 1|\mathbf{Y}_i^o, \mathbf{Y}_i^m, \mathbf{X}_i) = P(R_{ij} = 1|G_i)$$

with

$$P(R_{ij} = 1|G_i = W) < P(R_{ij} = 1|G_i = M).$$

This kind of MCAR is called MAR if the stronger definition of MCAR is used.

Missing at random (MAR)

$$P(R_{ij} = 1|\mathbf{Y}_i^o, \mathbf{Y}_i^m, \mathbf{X}_i) = P(R_{ij} = 1|\mathbf{Y}_i^o, \mathbf{X}_i).$$

for $i = 1, \ldots, N$, $j = 1, \ldots, n$.

- The probability of missingness $(P(R_{ij} = 0))$ is not related to the value that would have been observed if the value had not been missing, but depends on the observed values.
- The distribution of $\boldsymbol{Y}_i^m$ conditional on $\boldsymbol{Y}_i^o$ (and $\boldsymbol{X}_i$) is the same as the corresponding distribution among the complete cases.
- In practice, MAR is more frequent than MCAR!

Examples MAR

- Ethical considerations require that a patient is removed from the study if Y_{ij} falls outside a certain range of values (patient is not responding to the treatment).

- Creatinine level is too bad $\rightarrow$ patient is dialyzed in a different department/hospital.

- Respiratory problems in children $\rightarrow$ family moves to a place with better air quality.

. . . always assuming the decision is associated only with observed values $\boldsymbol{Y}_i^o$.

Not missing at random (NMAR)

$$P(R_{ij} = 1 | \mathbf{Y}_i^o, \mathbf{Y}_i^m, \mathbf{X}_i)$$

cannot be simplified as with MCAR or MAR, $i = 1, \ldots, N$, $j = 1, \ldots, n$.

- The probability of missingness ($P(R_{ij} = 0)$) depends on the observed as well as on the unobserved values.
- Also called informative missingness.
- NMAR is (unfortunately!) quite common.

Examples NMAR

- In a study on pain relief, patients with severe pain are less likely to answer the phone and give their current pain status.
- Heavy people hesitate to give their weight.
- Major respiratory problems $\rightarrow$ hospital!

Overview Chapter 11 - Missing values

11.1 Introduction

11.2 Three different missing value mechanisms

11.3 Special case: dropout

11.4 ML based methods and GEE for missing values

11.5 Overview of data analysis methods for missing values

11.6 Models for the dropout process

Dropout

- For intermittent missing values, the reason is often known, as subjects remain in the study $\rightarrow$ find out whether MCAR or MAR assumption is tenable $\rightarrow$ analysis of available data

- For dropout, we often have to suspect a relation between the dropout and the measurement process (MAR or NMAR).

Dropout: Possible reasons

- Other disease, death $\rightarrow$ MCAR only if unrelated to what is studied!
- Uncooperative patient $\rightarrow$ MCAR if unrelated to what is studied
- Ineffective therapy $\rightarrow$ MAR if decision based on Y_{ij}^o, otherwise NMAR
- Moving $\rightarrow$ MCAR, MAR or NMAR depending on reason
- Patient feeling too sick, which would be reflected in $\boldsymbol{Y}_i^m \rightarrow$ NMAR
- Unknown ("lost to follow-up": LOFU) $\rightarrow$??

Dropout: Graphical display

Examples:

- "Survival Curve"
- Individual curves grouped by dropout time

For MCAR, the history of y_{ij} values of people "about to drop out" should be the same (or conditional on $\boldsymbol{X}_i$) as that of those not dropping out.
$\rightarrow$ compare visually or for formal test see Diggle (1989).

Overview Chapter 11 - Missing values

11.1 Introduction

11.2 Three different missing value mechanisms

11.3 Special case: dropout

11.4 ML based methods and GEE for missing values

11.5 Overview of data analysis methods for missing values

11.6 Models for the dropout process

Likelihood-based inference and missing data

For likelihood-based inference, it is most important to distinguish between MCAR/MAR on the one hand, and NMAR on the other hand.

The joint density of $(\mathbf{Y}^o, \mathbf{Y}^m, \boldsymbol{R})$ is

$$f(\boldsymbol{y}^o, \boldsymbol{y}^m, \boldsymbol{r}|\boldsymbol{X}_i) = f(\boldsymbol{y}^o, \boldsymbol{y}^m|\boldsymbol{X}_i) f(\boldsymbol{r}|\boldsymbol{y}^o, \boldsymbol{y}^m, \boldsymbol{X}_i).$$

The joint density of the observable data then factors as

$$\begin{aligned} f(\boldsymbol{y}^o, \boldsymbol{r}|\boldsymbol{X}_i) &= \int f(\boldsymbol{y}^o, \boldsymbol{y}^m|\boldsymbol{X}_i) f(\boldsymbol{r}|\boldsymbol{y}^o, \boldsymbol{y}^m, \boldsymbol{X}_i) d\boldsymbol{y}^m \\ &\overset{MCAR/MAR}{=} \int f(\boldsymbol{y}^o, \boldsymbol{y}^m|\boldsymbol{X}_i) d\boldsymbol{y}^m f(\boldsymbol{r}|\boldsymbol{y}^o, \boldsymbol{X}_i) \\ &= f(\boldsymbol{y}^o|\boldsymbol{X}_i) f(\boldsymbol{r}|\boldsymbol{y}^o, \boldsymbol{X}_i). \end{aligned}$$

Likelihood-based inference and missing data

The log-likelihood then is

$$\log L = \log f(\boldsymbol{y}^o|\boldsymbol{X}_i) + \log f(\boldsymbol{r}|\boldsymbol{y}^o, \boldsymbol{X}_i).$$

It is maximized by maximizing the two terms separately. Since the second term contains no information about the distribution of $\boldsymbol{Y}^o$, we can ignore it for inference about $\boldsymbol{Y}^o$.

Thus, MCAR/MAR are sometimes jointly referred to as ignorable missingness.

Likelihood-based inference and missing data

However,

- "ignorability" depends on the likelihood being the basis for inference (and being correctly specified!). (Standard) GEE is only valid under the stronger assumption of MCAR.

- if $\log f(\boldsymbol{y}^o)$ and $\log f(\boldsymbol{r}|\boldsymbol{y}^o, \boldsymbol{X}_i)$ share parameters, ignoring $\log f(\boldsymbol{r}|\boldsymbol{y}^o, \boldsymbol{X}_i)$ will result in a loss of efficiency.

- this assumes that the distribution of $\boldsymbol{Y}^o$ is the target of inference.

  Example: A clinical trial for treatment of a life-threatening disease. Dropout is due to patients' death. Inference about the distribution of the survival time and the conditional distribution of $\boldsymbol{Y}^o$ given survival may be more meaningful than about the unconditional distribution of $\boldsymbol{Y}^o$.

GEE and missing data

- GEE is used for its consistency under misspecified covariance structures and without distributional assumptions if the mean model is correct.
- Score equation:

$$\sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}$$

- Only consistent for MCAR!
- Consider the probability p_{ij} of observing Y_{ij} conditional on the history $y_{i1}, \ldots, y_{i,j-1}$ and covariates.
- **Assumption**: Measurement y_{ij} is representative of missing values from subjects with similar history.

A variation of GEE

- Robins et al (1995) propose a weighted GEE for MAR, where each observed measurement gets the weight $1/p_{ij}$ (inverse probability weighting), upweighting measurements with small probabilities ($\mathbf{P}_i = \text{diag}(p_{ij})$):

$$\sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1} \mathbf{P}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}.$$

- The resulting estimator is consistent under certain conditions including that the p_{ij} are consistently estimated. $\rightarrow$ More suitable for large samples!

- It requires a parametric model for the p_{ij} (with the data providing sparse information on the dropout process), in a setting where a parametric model for the covariance structure is avoided.

Example 1 (Little, 2008)

Suppose $n_i = 2$ for all i, and we have the normal model

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \overset{iid}{\sim} \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \right) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

Suppose that Y_{i1} is observed for all N subjects, but Y_{i2} only for the first r (dropout). MAR assumption: missingness of Y_{i2} can depend on Y_{i1}, but conditional on Y_{i1}, it does not depend on Y_{i2}. The likelihood is

$$L_{ign}(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \boldsymbol{Y}^o) = \prod_{i=1}^{r} |\boldsymbol{\Sigma}|^{-1/2} \exp(-\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu})) \quad (11.1)$$

$$\times \prod_{i=r+1}^{N} \sigma_{11}^{-1/2} \exp(-\frac{1}{2}(Y_{i1} - \mu_1)^2 / \sigma_{11}).$$

Example 1

The likelihood can be factored into the marginal distribution of Y_{i1} and the conditional distribution of Y_{i2} given Y_{i1}. The ML estimates then are

$$\widehat{\mu}_1 = \frac{1}{N}\sum_{i=1}^{N} y_{i1} \qquad \widehat{\sigma}_{11} = \frac{1}{N}\sum_{i=1}^{N}(y_{i1} - \widehat{\mu}_1)^2$$

$$\widehat{\mu}_2 = \bar{y}_2 + \widehat{\beta}_{2|1}(\widehat{\mu}_1 - \bar{y}_1) \qquad \widehat{\sigma}_{22} = s_{22} + \widehat{\beta}^2_{2|1}(\widehat{\sigma}_{11} - s_{11})$$

$$\widehat{\sigma}_{12} = \widehat{\beta}_{2|1}\widehat{\sigma}_{11}$$

where $\bar{y}_j$ and s_{jk} are sample means and (co)variances from the complete cases and $\widehat{\beta}_{2|1} = s_{12}/s_{11}$ is the regression coefficient regressing Y_{i2} on Y_{i1} for the complete cases.

Example 1

- Large-sample standard errors can be based on the observed information matrix, or obtained based on bootstrapping the observed data.

- The ML estimate $\widehat{\mu}_2$ adjusts $\bar{y}_2$ using available information on the difference $(\widehat{\mu}_1 - \bar{y}_1)$ between averages based on all cases and on complete cases only, and information on the association between Y_{i1} and Y_{i2}.

- By contrast, calculating the empirical means and variances for the two time points would result in unadjusted estimates $\bar{y}_2$ and s_{22}. So would using GEE with a working independence assumption corresponding to ML estimation with $\sigma_{12} = 0$.

Overview Chapter 11 - Missing values

11.1 Introduction

11.2 Three different missing value mechanisms

11.3 Special case: dropout

11.4 ML based methods and GEE for missing values

11.5 Overview of data analysis methods for missing values

11.6 Models for the dropout process

Overview of data analysis methods for missing values

- Complete case analysis
- Available data analysis
- Imputation
- Selection models

Complete case analysis

- Non-completers are completely deleted.
- Inefficient, wasteful of data (in extreme cases, there are no subjects without missing values).
- Only valid for MCAR (rare in practice). For MAR or NMAR, this can introduce bias.
- Useful only if you are only interested in the completers, otherwise not recommended.

Available data analysis

- General term for methods that can analyse the available data with unequal n_i.
- More efficient than complete case analysis.
- Only valid for MCAR (rare in practice) or for MAR if likelihood-based methods are used.

Example 1 continued

$$(Y_{i1}, Y_{i2})^T \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

with Y_{i2} observed only for the first r subjects.

- A complete case analysis would be biased for MAR, yielding $\widehat{\mu}_j = \bar{y}_j$, $j = 1, 2$, based only on completers.

- An available case analysis for MAR is fine if ML with general $\boldsymbol{\Sigma}$ is used, but would be biased for independent mean estimation or GEE with incorrect working covariance (cf. p. 24-26).

Imputation methods

- Last value carried forward: if y_{ij} is the last observed value, y_{ik} is set to y_{ij} for subsequent missing values. Variations:
  - Estimate a time-trend and extrapolate.
  - Baseline value carried forward, worst value carried forward.

  Strong and often unrealistic assumptions! Data with less variability, over-optimistic standard errors. Not recommended.

- Methods which draw imputed $\mathbf{y}_i^m$ from $f(\mathbf{y}_i^m|\mathbf{y}_i^o, \mathbf{X}_i)$:
  - Propensity based methods
  - Predictive mean matching

  Subsequent analyses are valid under MAR or MCAR. Multiple imputation also ensures that uncertainty is properly accounted for.

Propensity based imputation

- These methods are based on a model for the dropout probability, such as e.g.

$$\log\left[\frac{P(D_i = k|D_i \geq k, Y_{i1}, \ldots, Y_{ik})}{P(D_i > k|D_i \geq k, Y_{i1}, \ldots, Y_{ik})}\right] = \theta_1 + \theta_2 Y_{ik-1}$$

  Which missing mechanism do we have here?

- Missing reponses are imputed based on responses of subjects with similar estimated dropout probability but who did not drop out.

Predictive mean matching

- Regression models for Y_{ik} based on $Y_{i1}, \ldots, Y_{ik-1}$:

$$E(Y_{ik}) = \gamma_1 + \gamma_2 Y_{i1} + \cdots + \gamma_k Y_{ik-1}$$

- Each model is estimated based on the subjects with $D_i > k$.
- This results in estimates $\widehat{\gamma}$ and $\widehat{\sigma}$ (error variance).
- To account for estimation uncertainty, values γ^* and σ^* are drawn from the distribution of $\widehat{\gamma}$ (and $\widehat{\sigma}$).
- This gives the imputed value

$$\gamma_1^* + \gamma_2^* Y_{i1} + \cdots + \gamma_k^* Y_{ik-1} + \sigma^* e_i,$$

with simulated $e_i \sim \mathcal{N}(0, 1)$. (Can be generalized to GLMs.)

Multiple imputation

- Each missing value is imputed by several (tyically $5 \leq m \leq 10$) values. Why is this useful?

- $\rightarrow$ m data sets are generated
  $\rightarrow$ m estimates $\widehat{\beta}^{(k)}$ and $\widehat{\mathsf{Cov}}(\widehat{\beta}^{(k)})$

- The result is (Rubin, 1987)

$$\overline{\beta} = \frac{1}{m} \sum_{k=1}^{m} \widehat{\beta}^{(k)}$$

$$\widehat{\mathsf{Cov}}(\overline{\beta}) = \frac{1}{m} \sum_{k=1}^{m} \widehat{\mathsf{Cov}}(\widehat{\beta}^{(k)}) + \left(1 + \frac{1}{m}\right) \frac{1}{m-1} \sum_{k=1}^{m} \left(\widehat{\beta}^{(k)} - \overline{\beta}\right) \left(\widehat{\beta}^{(k)} - \overline{\beta}\right)^T.$$

Further alternatives

- Weighting methods for MAR (cf. slide 23 for GEE), see e.g. Fitzmaurice et al. (2004), Chapter 14.

- The EM-algorithm for MAR, see e.g. Molenberghs & Verbeke (2005), Chapter 28.

The EM-algorithm is also an alternative if values are missing below the limit of detection / above a cut-off value (censoring).

Overview Chapter 11 - Missing values

11.1 Introduction

11.2 Three different missing value mechanisms

11.3 Special case: dropout

11.4 ML based methods and GEE for missing values

11.5 Overview of data analysis methods for missing values

11.6 Models for the dropout process

Models for dropout

- **Idea:** Joint modeling of the dropout mechanism and $\boldsymbol{Y}_i$
- Two important approaches: selection models and pattern mixture models
- Selection models are based on the factorization

$$f(\mathbf{y}_i, \mathbf{r}_i|\boldsymbol{X}_i, \boldsymbol{\theta}, \boldsymbol{\psi}) = f(\mathbf{y}_i|\boldsymbol{X}_i, \boldsymbol{\theta}) f(\mathbf{r}|\mathbf{y}_i, \boldsymbol{X}_i, \boldsymbol{\psi})$$

  with $f(\mathbf{r}_i|\mathbf{y}_i, \boldsymbol{X}_i, \boldsymbol{\psi}) = f(\mathbf{r}_i|\boldsymbol{X}_i, \boldsymbol{\psi})$ for MCAR and $f(\mathbf{r}_i|\mathbf{y}_i, \boldsymbol{X}_i, \boldsymbol{\psi}) = f(\mathbf{r}_i|\mathbf{y}_i^o, \boldsymbol{X}_i, \boldsymbol{\psi})$ for MAR.

- Pattern mixture models are based on the factorization

$$f(\mathbf{y}_i, \mathbf{r}_i|\boldsymbol{X}_i, \boldsymbol{\nu}, \boldsymbol{\delta}) = f(\mathbf{r}_i|\boldsymbol{X}_i, \boldsymbol{\delta}) f(\mathbf{y}_i|\mathbf{r}_i, \boldsymbol{X}_i, \boldsymbol{\nu}).$$

Example 1 continued

Consider a selection model for NMAR dropout:

$$(Y_{i1}, Y_{i2})^T \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$
$$(R_{i2}|Y_{i1}, Y_{i2}) \sim \text{Bernoulli}(\pi_i)$$
$$\text{logit}(\pi_i) = \psi_0 + \psi_1 Y_{i1} + \psi_2 Y_{i2}.$$

The likelihood is

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\psi}|\boldsymbol{R}, \boldsymbol{Y}^o) = \prod_{i=1}^{r} |\boldsymbol{\Sigma}|^{-1/2} \exp(-\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}))\pi_i(\boldsymbol{\psi})$$
$$\times \prod_{i=r+1}^{N} \int |\boldsymbol{\Sigma}|^{-1/2} \exp(-\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}))(1 - \pi_i(\boldsymbol{\psi}))dY_{i2}.$$

Example 1 continued

- Maximization requires an iterative algorithm such as the EM algorithm
- The model is weakly identified, and identification is strongly depending on the model assumptions.
- Thus, it is preferred to either make additional assumptions such as $\psi_1 = 0$ or $\psi_2 = 0$, or to conduct a sensitivity analysis for a range of plausible $\boldsymbol{\psi}$.
- For $\psi_2 = 0$ (MAR), the likelihood reduces to

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\psi} | \boldsymbol{R}, \boldsymbol{Y}^o) = L_{ign}(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \boldsymbol{Y}^o) \prod_{i=1}^{r} \pi_i(\boldsymbol{\psi}) \prod_{i=r+1}^{N} (1 - \pi_i(\boldsymbol{\psi})),$$

  where $L_{ign}(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \boldsymbol{Y}^o)$ is given by (11.1), and ML estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ can be based on the ignorable likelihood, as discussed in example 1.

Overview applicability of methods

	MCAR	MAR	NMAR
Expected value, variance	yes (or conditional on $\boldsymbol{X}_i$)	no	no
Available case analysis	yes	no (GEE)/yes (ML)	no
Complete case analysis	yes, but inefficient	no	no
GEE	yes	no, or weighted GEE	no
ML methods	yes	yes	no
(Multiple) imputation from $f(\boldsymbol{y}_i^m \vert \boldsymbol{y}_i^o, \boldsymbol{X}_i)$	yes	yes	no
Selection / pattern mixture models	yes	yes	yes

Discussion

- ML (or Bayesian) inference for ignorable missingness is similar to corresponding complete data analyses. However, randomness (MAR) of missings is an assumption which cannot be verified from the observed data.

- Non-ignorable models are more challenging, have problems with lack of identifiability and require assumptions about the missing data mechanism, e.g. a pametric model for $\boldsymbol{R}_i$ given $\boldsymbol{Y}_i$ and $\boldsymbol{X}_i$ in selection models.

- Oftentimes, especially in potential NMAR cases, a sensitivity analysis under different assumptions is the most sensible alternative to make the dependence of results on assumptions transparent.

- If covariate values are also missing, additional work is required, with multiple imputation being one option.

- Read more e.g. in Diggle et al (2002), Molenberghs & Verbeke (2005) or Fitzmaurice et al. (2008).