Concepts of missing data

Samiran Sinha

Texas A&M University

sinha@stat.tamu.edu

November 29, 2017

Longitudinal dataset

- Consider the cervical dystonia dataset given at `http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets`
- Read the `.dta` data file in `R`

Example

```
a=read.dta("cdystonia.dta")
head(a)
  week site id treat age sex twstrs
1    0    1  1 5000U  65   F     32
2    2    1  1 5000U  65   F     30
3    4    1  1 5000U  65   F     24
4    8    1  1 5000U  65   F     37
5   12    1  1 5000U  65   F     39
6   16    1  1 5000U  65   F     36
```

Longitudinal dataset

- The response variable `twstrs` is supposed to be be available at the baseline (0), 2, 4, 8, 12, and 16 weeks for each of the subjects.
- Not all subjects have all 6 measurements, the type of measurements can be classified as follows: for some subjects only 4 responses were recorded, for some 5 responses, and for the rest all 6 responses were recorded.

A model

- A : placebo, B : experimental treatment with 5000U, C : experimental treatment with 10000U
- X_i : treatment, either A, B, or C
- t_j : the jth time point
- Suppose the model for the response $Y_{i,j}$ is

$$
\begin{aligned}
Y_{i,j} &= \beta_0 + \beta_1 I(X_i = B) + \beta_2 I(X_i = C) + \delta I(\mathrm{Gender}_i = \mathrm{Male}) \\
&\quad +\gamma_0 t_j + \gamma_1 t I(X_i = B) + \gamma_2 t I(X_i = C) \\
&\quad a_i + e_{i,j},
\end{aligned}
$$

- β_1, β_2: treatment related main effect parameters; γ_0: time slope; γ_1, γ_2: treatment versus time interaction parameters
- a_i subject specific random effect, assume a_i iid $\mathrm{Normal}(0, \sigma_a^2)$, $e_{i,j}$ pure noise assume $e_{i,j}$ iid $\mathrm{Normal}(0, \sigma_e^2)$
- I kept the model simple, but if you want, you may include more iteration terms in the model.

Mechanism

- What is missing here? The response $Y_{i,j}$ is missing, the predictors are completely observed.
- Why are some $Y_{i,j}$'s missing? Does the chance of missing values depends on the latent $Y_{i,j}$ value itself?
- Define a new indicator variable $R_{i,j} = 1$ if $Y_{i,j}$ is observed and 0 otherwise.
- Missing completely at random (MCAR): missingness mechanism does not depend on any variable. Here $\text{pr}(R_{i,j} = 1)$ is a constant does not depend on $Y_{i,j}$ or on any predictor variable.
- Missing at random (MAR): missingness mechanism does depend only on the observed values of the variables. A special case is here $\text{pr}(R_{i,j} = 1)$ may depend on treatment, time, gender, and site, but not on the response.
- Missing not at random (MNAR): missingness mechanism may depend on all the variables, including the missing values of the response, that means here $\text{pr}(R_{i,j} = 1)$ may depend on treatment, time, gender, site, and response. This is also called non-ignorable mechanism.

Mechanism verification

- MCAR: Chen & Little (1999) proposed a test to verify this assumption
- MAR: Direct verification of this assumption is not possible. However, under some scenarios (constraints) it can be tested (Potthoff et al, 2006; Sinha et al., 2014).
- MNAR: This cannot be verified, therefore, people perform sensitivity analysis when MNAR is suspected.

Mechanism

- Consider a simple scenario of MAR, that $\mathrm{pr}(R_{i,j} = 1)$ does not depend on Y values at all. Then consider the conditional expectation of the response only among the observed responses given all the predictors (treatment, site, gender etc),

$$
\begin{aligned}
E(Y|R=1,Z) &= \int r f_y(r|R=1,Z)dr \\
&= \int r \frac{\mathrm{pr}(R=1|r,Z)f_y(r|Z)dr}{\int \mathrm{pr}(R=1|u,Z)f_y(u|Z)du} \\
&= \int r \frac{\mathrm{pr}(R=1|Z)f_y(r|Z)dr}{\int \mathrm{pr}(R=1|Z)f_y(u|Z)du} \\
&= \int r f_y(r|Z)dr \\
&= E(Y|Z)
\end{aligned}
$$

This relation implies that for estimating the conditional expectation of Y given Z it is okay to consider only the observed Y-values.

Mechanism

- Following the similar idea, we can infer for MCAR scenario, it is okay to make inference on $E(Y|Z)$ based on the available data only.
- Consider another scenario of MAR where $\text{pr}(R_{i,j}=1)$ does depend only on the observed values of the response, but not on the $Y_{i,j}$th value. Then

  $$E(Y|R=1,Z) \quad \neq \quad E(Y|Z),$$

  so analysis purely based on the observed responses may not characterize the population of Y given Z.
- Suppose that missingness is influenced by some subject specific characteristics that are related with the random subject specific term a_i. This is a type of MNAR. In this case,

  $$E(Y|R=1,Z) \quad \neq \quad E(Y|Z),$$

Missing pattern

- For every subject i, if missingness of $Y_{i,j}$ implies all $Y_{i,k}$'s are missing for $k > j$, then the pattern is called monotonic. That means, once a subject drops out from the study, he/she never returns again.
- If the missing pattern is not monotonic, we call it non-monotonic pattern.
- A careful study of the cervical dystonia data reveals that there is a non-monotone pattern of missing data.

Multiple imputation method

- What should we do, if we have MCAR, MAR or MNAR data?
- For MCAR or MAR, to increase the efficiency of the estimator, we can follow the multiple imputation method.
- Missing values are filled-in with some plausible values.
- This process is repeated for multiple times, creating several complete datasets. Usually 5-10 datasets are created by this mechanism.
- Each dataset is analyzed by a standard method, and then we combine the results from each dataset.
- The main concern is about the imputation model that is used to generate plausible values for the missing entries.

Multiple imputation method

- For our example, for the t_jth time point we could fit a linear regression model

  $$Y_{i,j} = \alpha_0^{(j)} + \alpha_1^{(j)} I(X_i = B) + \alpha_2^{(j)} I(X_i = C) + \alpha_3^{(j)} I(\mathrm{Gender}_i = \mathrm{Male}) + u_i^{(j)}$$

  based on the observed data.
- Obtain $\widehat{\boldsymbol{\alpha}}^{(j)} = (\widehat{\alpha}_0^{(j)}, \widehat{\alpha}_1^{(j)}, \widehat{\alpha}_2^{(j)}, \widehat{\alpha}_3^{(j)})^T$ and asymptotic variance $\Sigma^{(j)}$.
- Simulate $\boldsymbol{\alpha}_1^{(j)}, \ldots, \boldsymbol{\alpha}_m^{(j)}$ from $\mathrm{Normal}(\widehat{\boldsymbol{\alpha}}^{(j)}, \Sigma^{(j)})$.
- Replace the missing $Y_{i,j}$'s by
  $(1, I(X_i = B), I(X_i = C), I(\mathrm{Gender}_i = \mathrm{Male}))^T \boldsymbol{\alpha}_l^{(j)}$, $l = 1, \ldots, m$.
- That way create m imputed datasets (m *complete* datasets).
- Analyze each dataset by the standard method, and then combine the results in the end.
- Note that imputation model is different from the actual model.

Hot deck imputation method

- Missing values are filled-in with the observed responses of *similar* units.
- If there are multiple similar units, then a unit is chosen randomly from the pool of similar units.
- Then the observed response of the selected unit is used to replace the missing value.
- Once the missing values are filled-in, then you proceed with the analysis of the
- For further reference see Andridge and Little (2010).

Sleep data

- Consider the following dataset in the `VIM` package
- A description of the dataset is given in Sleep in mammals: Ecological and constitutional correlates. *Science*, **194**, 732–734, (1976).

Example

```
 library(VIM)
  head(sleep)
    BodyWgt BrainWgt NonD Dream Sleep  Span  Gest Pred Exp Danger
1  6654.000  5712.00   NA    NA   3.3  38.6 645.0    3   5      3
2     1.000     6.60  6.3   2.0   8.3   4.5  42.0    3   1      3
3     3.385    44.50   NA    NA  12.5  14.0  60.0    1   1      1
4     0.920     5.70   NA    NA  16.5    NA  25.0    5   2      3
5  2547.000  4603.00  2.1   1.8   3.9  69.0 624.0    3   5      4
```

Continues

- Data on 62 species
- Bodyweight in kilograms
- Brainweight in grams
- Gest: gestation time in days
- Sleep (dream) in hours per day
- pred: severity of predation in a scale from 1 to 5 (1 means minimum predation and 5 maximum predation)
- exp: sleep exposure (1 for those who sleep in well protected area, 5 least protected area)
- danger: overal estimate of predatory danger

complete cases

Example

```
 summary(sleep)
 dim(sleep)
[1] 62 10
 newdata=sleep[complete.cases(sleep),]
 dim(newdata)
[1] 42 10
## Show cases with at least one missing value in 10 variables
sleep[!complete.cases(sleep),]
# the following code returns the percentage (frequency) of missing
# observations for each of the variables prop=T
# (prop=F)
aggr(sleep, prop = T, numbers = T)
aggr(sleep, prop = F, numbers = T)
```

Fitting the imputation model (fully conditional models, chain rule)

- Response Y: Sleep, all other variables are predictors, $X_1, \ldots, X_9$
- Missing values are in predictors as well as in Y
- X_4: span, X_3: gest, X_2: dream, X_1 : nonD
- No missig values in X_5 : bodyweight, X_6 : brainweight, X_7 : pred, X_8 : exp, X_9 : danger
- Fit a model $f(Y|X_1, \cdots, X_9)$ using the data where $(Y, X_1, \ldots, X_9)$ are all observed
- Fit a model $f(X_4|Y, X_1, X_2, X_3, X_5, \cdots, X_9)$ using the data where $(Y, X_1, \ldots, X_9)$ are all observed
- Fit a model $f(X_3|Y, X_1, X_2, X_4, \cdots, X_9)$ using the data where $(Y, X_1, \ldots, X_9)$ are all observed
- Fit a model $f(X_2|Y, X_1, X_3, \cdots, X_9)$ using the data where $(Y, X_1, \ldots, X_9)$ are all observed
- Fit a model $f(X_1|Y, X_2, \cdots, X_9)$ using the data where $(Y, X_1, \ldots, X_9)$ are all observed

Predictive mean matching method (need to re-word)

1 For ease of presentation of the concept, assume that only Y, X_1, X_2 have missing values, the idea can be extended for missing values in more predictors.

2 For cases with no missing data, estimate a linear regression of Y on $X_1, \dots, X_9$(conditioning variables), producing a set of coefficients $\boldsymbol{\theta}_y$, estimate a linear regression of X_1 on $Y, X_2, \dots, X_9$(conditioning variables), producing a set of coefficients $\boldsymbol{\theta}_1$, and estimate a linear regression of X_2 on $Y, X_1, X_3 \dots, X_9$(conditioning variables), producing a set of coefficients $\boldsymbol{\theta}_2$

3 Make a random draw from the 'posterior predictive distribution' of $\boldsymbol{\theta}_y$, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ separately, producing a new set of coefficients $\boldsymbol{\theta}_y^*$, $\boldsymbol{\theta}_1^*$ and $\boldsymbol{\theta}_2^*$. Usually this would be a random draw from a multivariate normal distribution with mean $\widehat{\boldsymbol{\theta}}$ and the estimated covariance matrix of $\widehat{\boldsymbol{\theta}}$ (with an additional random draw for the residual variance). This step is necessary to produce data imputing observations from their predictive density.

Predictive mean matching method

4. Using $\boldsymbol{\theta}^*$, generate predicted values for Y, X_1, X_2 for all cases, irrespective of whether a case contain a missing of value or not. This generation is done through iterations, and with each iteration each variable (such Y, X_1, or X_2) simulated conditional on the recent simulated values of the other variables, while completely observed variables are always fixed to their observed values.
5. For each case with missing Y, identify a set of cases with observed Y whose predicted values for (Y, X_1, X_2) are close to the predicted value for the case with missing data. From among those close cases (based on some distance measure), one case is randomly chosen, and use its observed value to replace the missing value. Repeat this step, for X_1, and X_2.
6. Repeat steps 3-5 to get several imputed datasets.

Mice based on fully conditional specification

Example

```
library(mice)
 # creates imputed datasets, m denotes the number of imputed datasets
 # we want
 mi.sleep <- mice(sleep, seed = 1234, m=6, printFlag = FALSE)
 mydata1=complete(mi.sleep, action = 1) # this is the first imputed
 # dataset
```

Sleep data

Example

```
library(mice)
 # creates imputed datasets, m denotes the number of imputed
 #  datasets we want
mi.sleep <- mice(sleep, seed = 1234, m=6, printFlag = FALSE)
store.coef=NULL
store.var=diag(10)*0;
for(k in 1:6){
mydataimpu=complete(mi.sleep, action = k)
out=lm(Sleep~ BodyWgt+ BrainWgt+ NonD+ Dream + Span+ Gest+
Pred+ Exp+ Danger, data=mydataimpu)
store.coef=rbind(store.coef, out$coefficient)
store.var=store.var+(summary(out)$cov.unscaled)*(
(summary(out)$sigma)^2)
}
```

Sleep data– estimates after imputation

$\widehat{\beta} = \sum_{l=1}^{m} \widehat{\beta}_l / m$
$\text{Var}(\widehat{\beta}) = \sum_{l=1}^{m} \text{Var}(\widehat{\beta}_l)/m + \{(m+1)/m\} \sum_{l=1}^{m} (\widehat{\beta}_l - \widehat{\beta})^2/(m-1)$ The above formula provides a conservative estimate of $\text{Var}(\widehat{\beta})$.

Example

```
final.estimate=apply(store.coef, 2, mean)
final.var=  store.var/m+var(store.coef)
```

Variance of the multiple imputation estimator

- $\text{Var}(\widehat{\beta}) = \sum_{l=1}^{m} \text{Var}(\widehat{\beta}_l)/m + \{(m+1)/m\} \sum_{l=1}^{m} (\widehat{\beta}_l - \widehat{\beta})^2/(m-1)$
- Note that the second term is zero if there is no missing data– therefore $\theta \equiv \sum_{l=1}^{m} (\widehat{\beta}_l - \widehat{\beta})^2/(m-1)/\sum_{l=1}^{m} \text{Var}(\widehat{\beta}_l)/m$ is a measure of missing information.
- The appropriate confidence interval for the parameter is then

$$\widehat{\beta} \pm t_{df}\sqrt{\text{Var}(\widehat{\beta})},$$

where $df = (m-1)(1+1/r^2)$, $r = (1+1/m)\theta$ (Rubin, 1987)

Sleep data– estimates after imputation

If any of the variables that has missing value is of binary (0/1) type, then we can use `logreg` option in the `mice` package.

Example

```
 myx=ifelse(sleep$NonD<5, 0, 1) # creates a binary variable out of
 # NonD
newdata=data.frame(sleep, myx)
newdata$myx<-as.factor(newdata$myx)
mi.sleep <- mice(newdata[, -3], seed = 1234, m=5, printFlag =
FALSE, defaultMethod=c("pmm", "pmm", "pmm", "pmm", "logreg"))
```

Sleep data– estimates after imputation

Suppose that we don't want to use gest as the predictor of other variables. Then we will follow the following code.

Example

```
out0 = mice(newdata[, -3], maxit=0)
meth = out0$method
predM = out0$predictorMatrix
predM[, c("Gest")]=0

meth[c("Dream")]="norm"
meth[c("Sleep")]="norm"
meth[c("Span")]="norm"
meth[c("Gest")]="norm"
meth[c("myx")]="logreg"
out2 = mice(newdata[, -3], seed=1234, method=meth, predictorMatrix=predM,
 m=5, printFlag=FALSE)
```

Imputation using joint model

- Response Y: Sleep, all other variables are predictors, $X_1, \ldots, X_9$
- Missing values are in predictors as well as in Y
- To have a concrete idea suppose that there are two predictors, X_1 and X_2, and both of them have missing values.
- For imputing X_1 and X_2 for a particular observation, we shall use the model $f(X_1, X_2|Y)$ conditional on Y
- Note that $f(X_1, X_2|Y) = f(X_1|X_2, Y)f(X_2|Y)$. So, first simulate X_2 given Y and then simulate X_1 conditional on X_2 and Y.
- The question is if Y is also missing for that subject, what should we do? Then, first simulate Y from a model $f(Y)$ (not conditional on X_1 or X_2), then given Y, sample X_2 from $f(X_2|Y)$, then sample X_1 from $f(X_1|X_2, Y)$.

Fitting the imputation model (joint modelling)

- X_4: span, X_3: gest, X_2: dream, X_1 : nonD
- No missig values in X_5 : bodyweight, X_6 : brainweight, X_7 : pred, X_8 : exp, X_9 : danger
- Fit a model $f(Y|X_5, \cdots, X_9)$ using the data where $(Y, X_5, \ldots, X_9)$ are all observed
- Fit a model $f(X_4|Y, X_5, \cdots, X_9)$ using the data where $(X_4, Y, X_5, \ldots, X_9)$ are all observed
- Fit a model $f(X_3|X_4, Y, X_5, \cdots, X_9)$ using the data where $(X_3, X_4, Y, X_5, \ldots, X_9)$ are all observed
- Fit a model $f(X_2|X_3, X_4, Y, X_5, \cdots, X_9)$ using the data where $(X_2, X_3, X_4, Y, X_5, \ldots, X_9)$ are all observed
- Fit a model $f(X_1|X_2, X_3, X_4, Y, X_5, \cdots, X_9)$ using the data where $(X_1, \ldots, X_4, Y, X_5, \ldots, X_9)$ are all observed

Imputing missing values

- For imputing the NonD (X_1) and Dream (X_2) values for the first subject in the sleep data:
  - Dream is simulated from $f(X_2|X_3 = 645, X_4 = 38.6, Y = 3.3, X_5 = 6654, \cdots, X_9 = 3)$, call the simulated value of Dream as X_2^*
  - Then NonD is simulated from $f(X_1|X_2 = X_2^*, X_3 = 645, X_4 = 38.6, Y = 3.3, X_5 = 6654, \cdots, X_9 = 3)$
- Repeat this process for each of 62 species (rows) whichever contains any missing entry
- For multiple imputation we repeat the imputation procedure *m* times, resulting in *m* imputed datasets

Example

```
 library(VIM)
  head(sleep)
    BodyWgt BrainWgt NonD Dream Sleep  Span  Gest Pred Exp Danger
1  6654.000  5712.00   NA    NA   3.3  38.6 645.0    3   5      3
```

Likelihood based approach

- Suppose that the data are (Y, X, Z, R)
  - Y : response
  - X : partially missing
  - $R = 1$ if X is observed and 0 otherwise
  - Z : always observed
- Goal is estimation of the model parameter θ of $f(Y|X, Z, \theta)$
- For likelihood based analysis we need a model for $f(X|Z, \gamma)$
- If the missingness is MAR or MCAR, then we do not need a probability model R, if the missingness is NMAR, then we need a model for R

Likelihood based approach for MAR or MCAR scenario

$$\mathcal{L} = \prod_{R_i=1} f(Y_i|X_i, Z_i, \theta) f(X_i|Z_i, \gamma) \prod_{R_i=0} \int f(Y_i|X_i, Z_i, \theta) f(X_i|Z_i, \gamma) dX_i$$

The parameters θ and γ are determined by maximizing $\mathcal{L}$. Usually EM algorithm is used if the computation of θ and γ are straight forward when X is always observed

Likelihood based approach for MNAR scenario

First we model probability of $R = 1$ given X, Y, Z in terms of a parametric model, say, $\text{pr}(R = 1|X, Y, Z, \eta)$

$$\begin{aligned}\mathcal{L} &= \prod_{R_i=1} \text{pr}(R = 1|X_i, Y_i, Z_i, \eta) f(Y_i|X_i, Z_i, \theta) f(X_i|Z_i, \gamma) \\ &\quad \times \prod_{R_i=0} \int \{1 - \text{pr}(R = 1|X_i, Y_i, Z_i, \eta)\} f(Y_i|X_i, Z_i, \theta) f(X_i|Z_i, \gamma) dX_i\end{aligned}$$

- One choice for the probability for R is logistic

$$\text{pr}(R = 1|X, Y, Z, \eta) = \frac{\exp(\eta_0 + \eta_1 Y + \eta_2 Z + \eta_3 X)}{1 + \exp(\eta_0 + \eta_1 Y + \eta_2 Z + \eta_3 X)}$$

- Write the negative of log-likelihood function, and may then use the `optim` function to minimize the objective function.
- Often η_3 you may face non-convergence issue, then you fix η_3 to different grid of values, and maximize it– this is called sensitivity analysis.

Likelihood based approach X and Y both missing

- Suppose that the data are (Y, X, Z, R_x, R_y)
    - Y : response, partially missing
    - X : predictor, partially missing
    - $R_x = 1$ if X is observed and 0 otherwise
    - $R_y = 1$ if Y is observed and 0 otherwise
    - Z : always observed
- Goal is estimation of the model parameter θ of $f(Y|X, Z, \theta)$
- For likelihood based analysis we need a model for $f(X|Z, \gamma)$

Formulation

- R_x may only depend on Z, R_y may only depend on Z
- There may be dependence between R_x and R_y

$$
\begin{aligned}
\mathcal{L} \;=\; & \prod_{R_{x,i}=1, R_{y,i}=1} f(Y_i|X_i, Z_i, \theta) f(X_i|Z_i, \gamma) \\
& \times \prod_{R_{x,i}=0, R_{y,i}=1} \int f(Y_i|X_i, Z_i, \theta) f(X_i|Z_i, \gamma) dX_i \\
& \times \prod_{R_{x,i}=1, R_{y,i}=0} \underbrace{\int f(Y_i|X_i, Z_i, \theta) dY_i}_{1} f(X_i|Z_i, \gamma) \\
& \times \prod_{R_{x,i}=0, R_{y,i}=0} \underbrace{\int f(Y_i|X_i, Z_i, \theta) f(X_i|Z_i, \gamma) dX_i dY_i}_{1}
\end{aligned}
$$

Formulation

Therefore,

$$\begin{aligned}
\mathcal{L} &= \prod_{R_{x,i}=1,R_{y,i}=1} f(Y_i|X_i, Z_i, \theta) f(X_i|Z_i, \gamma) \\
&\quad \times \prod_{R_{x,i}=0,R_{y,i}=1} \int f(Y_i|X_i, Z_i, \theta) f(X_i|Z_i, \gamma) dX_i \\
&\quad \times \prod_{R_{x,i}=1,R_{y,i}=0} f(X_i|Z_i, \gamma).
\end{aligned}$$

Complex scenario

- R_x may depend on Z and Y
- R_y may depend on X and Z
- R_x and R_y are independent conditional on X, Y, Z
- Two selection models are $\text{pr}(R_x = 1|Y, Z, \eta_x)$ and $\text{pr}(R_y = 1|X, Z, \eta_y)$
- This mechanism falls under MNAR mechanism

Complex scenario

The likelihood function for this complex scenario

$$
\begin{aligned}
& \mathcal{L}(\theta, \gamma, \eta_x, \eta_y) \\
= & \prod_{R_{x,i}=1, R_{y,i}=1} \mathrm{pr}(R_x = 1 | Y_i, Z_i, \eta_x) \mathrm{pr}(R_y = 1 | X_i, Z_i, \eta_y) f(Y_i | X_i, Z_i, \theta) f(X_i | Z_i, \gamma) \\
& \times \prod_{R_{x,i}=0, R_{y,i}=1} \mathrm{pr}(R_x = 0 | Y_i, Z_i, \eta_x) \int \mathrm{pr}(R_y = 1 | X_i, Z_i, \eta_y) f(Y_i | X_i, Z_i, \theta) f(X_i | Z_i, \gamma) dX_i \\
& \times \prod_{R_{x,i}=1, R_{y,i}=0} \mathrm{pr}(R_y = 0 | X_i, Z_i, \eta_y) f(X_i | Z_i, \gamma) \int \mathrm{pr}(R_x = 1 | Y_i, Z_i, \eta_x) f(Y_i | X_i, Z_i, \theta) dY_i \\
& \times \prod_{R_{x,i}=0, R_{y,i}=0} \int pr(R_x = 0 | Y_i, Z_i, \eta_x) \mathrm{pr}(R_y = 0 | X_i, Z_i, \eta_y) f(Y_i | X_i, Z_i, \theta) f(X_i | Z_i, \gamma) dX_i dY_i
\end{aligned}
$$

Will not consider other complex scenarios

References

Andridge, RA. & Little, RJA. (2010). A review of hot deck imputation for survey non-response. *International Statistical Review*, 78, 40–64.

Chen, HY. & Little, RJA. (1999). A test of missing completely at random for generalized estimating equations with missing data. *Biometrika*, 86, 1–13.

Potthoff, RF., Tudor, GE., Pieper, KS., & Hasselblad, V. (2006). Can one assess whether missing data are missing at random in medical studies? *Statistical Methods in Medical Research*, 15, 213–234.

Rubin, D. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Sinha, S., Saha, SK., & Wang, S. (2014). Semiparametric approach for non-monotone missing covariates in a parametric regression model. *Biometrics*, 70, 299–311.