

Multiple Imputation using Chained Equations (MICE)



Why missing data is a problem in NIS studies?

- For registries, Non-Interventional Studies data are collected on procedures and assessments, which are considered as part of standard practice.
 - Leads to a lot of missing data.
 - Even if 10-15% is missing per covariate, easily leads to only 40-50% of complete cases.
- Loss of power
- Alter important risk/benefit relationships
- Biased standard errors
- Inefficient estimates
- Standard ways of dealing with missing values, such as single imputation or complete case analysis (CCA) are generally inappropriate due to the loss of precision and risk of bias.
- Multiple imputation is the most flexible solution to missing data. When we have missing covariates, those observations won't contribute to normal MI analysis.

Idea of MICE

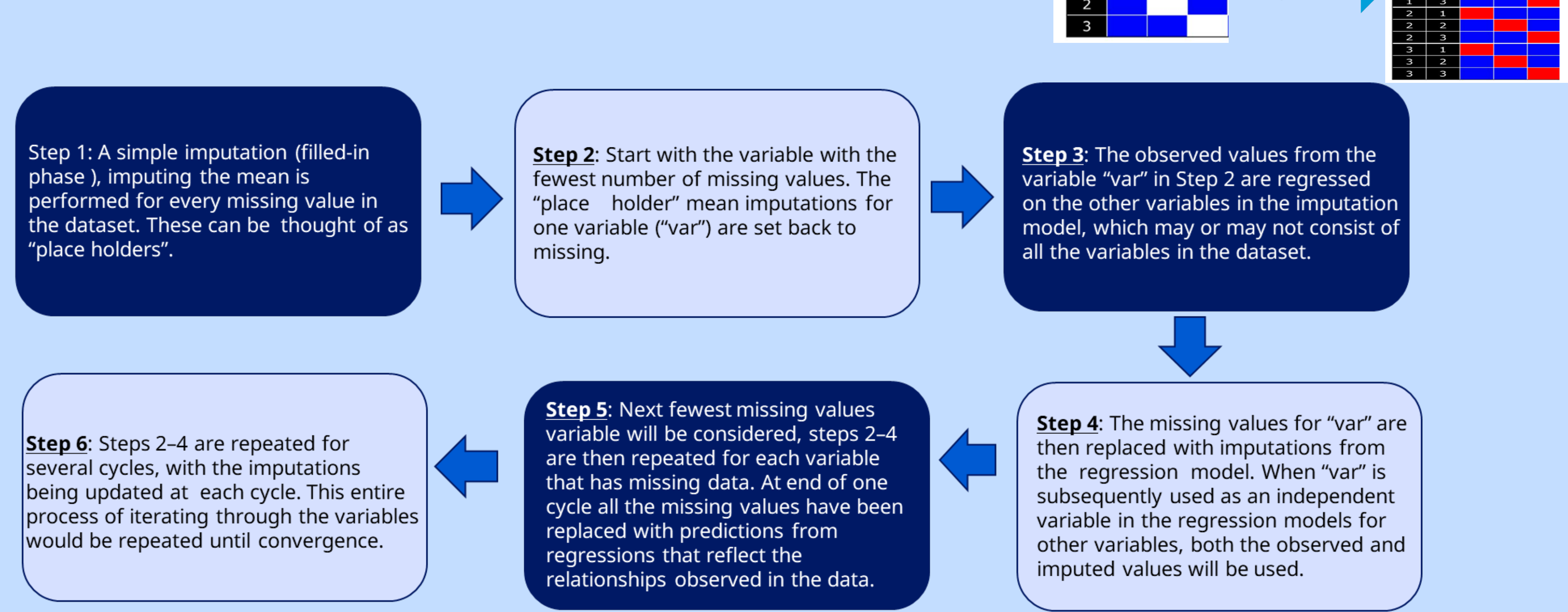
(Raghunathan et al., 2001; Van Buuren, 2007)

- In MICE each incomplete variable has its own imputation model (i.e. set of other variables used to predict it). Covariates are both continuous and categorical. Can be linear/logistic/ ordered logistic regression, tailored to variable type.
- Multiple imputation of missing data using the "chained equations/sequential regression /FCS/mice" method are all refer to the same approach.
- The Method available in SAS imputes missing variables using the fully conditional specification method (FCS) which does not assume a joint distribution but instead uses a separate conditional distribution for each imputed variable.

Assumption :

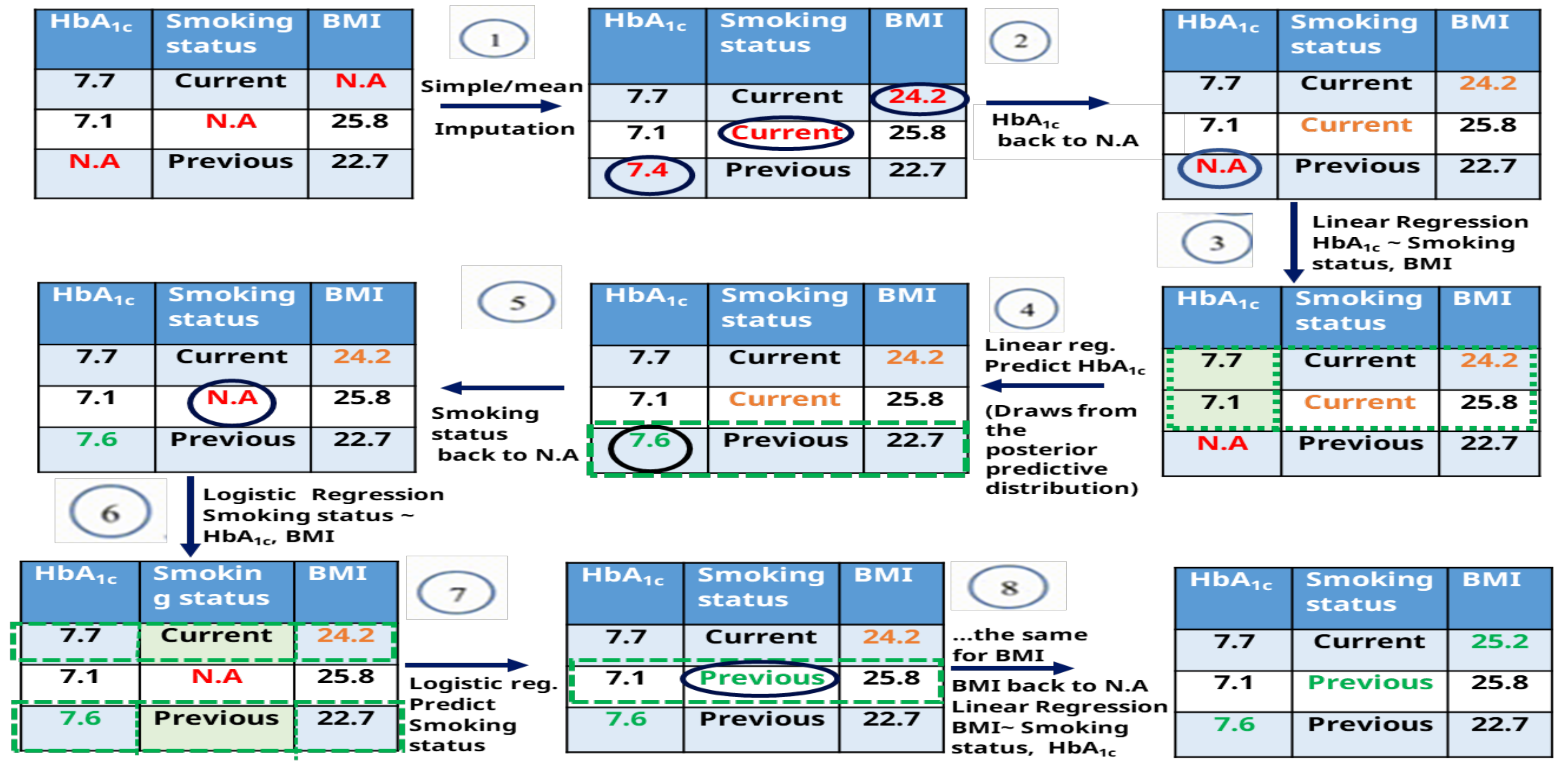
Given the variables used in the imputation procedure, the missing data are **Missing At Random (MAR)**, which means that the probability that a value is missing depends only on observed values and not on unobserved values.

MICE steps



The observed data and the final set of imputed values would then constitute one "complete" data set. Once the designated number of cycles has been completed, the **entire imputation process is repeated to generate multiple imputed datasets**. The observed data will be the same across the imputed datasets; only the values that had originally been missing will be imputed.

Multiple Imputation by Chained Equations (MICE) – single iteration



Application of FCS methods in SAS

Binary/categorical variables

- Discriminant function
- Logistic regression

Continuous variables

- Linear regression
- Predictive mean matching

By default the Discriminant function and regression method are used for imputation.

Imputation Methods	
DISCRIM	Specifies the discriminant function method
LOGISTIC	Specifies the logistic regression method
REG	Specifies the regression method
REGPMM	Specifies the predictive mean matching method

```
Proc mi data= inputdata nimpute=100 out=fcs ;
class treatment gender; fcs plots=trace(mean std);
var HbA1c smoking_status BMI treatment gender;
fcs discrim(smoking_status /classeffects = include) nbiter =100 ;
run;
```

```
Proc mi data= inputdata nimpute=100 out=fcs ;
class treatment gender; fcs plots=trace(mean std);
var HbA1c smoking_status BMI treatment gender;
fcs logistic(smoking_status/link=glogit) nbiter =100 ;
run;
```

```
Proc mi data= inputdata nimpute=100 out=fcs ;
class treatment gender; fcs plots=trace(mean std);
var HbA1c smoking_status BMI treatment gender;
fcs logistic (smoking_status/link=glogit) regpmm (HbA1c BMI)
nbiter =100 ; run;
```

```
Proc mi data= inputdata nimpute=100 out=fcs ;
class treatment gender; fcs plots=trace(mean std);
var HbA1c smoking_status BMI treatment gender;
fcs logistic(smoking_status = HbA1c treatment /link=glogit)
regpmm(HbA1c BMI) nbiter =100 ;
run;
```

Limitations of MICE

- MICE lacks the theoretical justification as compared to some other well-developed imputation approaches like MCMC.
- Specifying different distributions can lead to **slow convergence or non-convergence** of the imputation model.
- For MICE, each conditional density must be specified separately, so substantial **modeling effort** can be needed for data sets with many variables.
- MICE assumes of MAR. For missing data which is MNAR, new methods generating MIs under MNAR model will be required for handling such kind of missing data.