



Daria Shamaida, Quartesian Clinical Research, Kyiv, Ukraine

**MI FOR MI,
OR HOW TO HANDLE MISSING INFORMATION
WITH MULTIPLE IMPUTATION**



Agenda

- Two main missing data patterns
- Missing data mechanisms
- Strategies for dealing with missing data
- Why MI?
- Steps of performing multiple imputation
- Imputation methods, depending on missing data pattern and variable type
- Analysis Endpoints
 - Step 0 (before MI) – testing missing pattern
 - Step 1 - Multiple imputation: MAR
 - Step 1 - Multiple imputation: MNAR
 - Step 2 - Mixed effects model with repeated measures (MMRM)
 - Step 3 – Combining the results, using PROC MIANALYZE
 - Results of the Primary and Sensitivity Analyses
- Conclusion

Two main missing data patterns

- Monotone missing data pattern

It occurs when the value of the variable is missed not only in the specific time point for the particular subject, but in all subsequent time occasions for this individual.

Obs	Variable 1	Variable 2	Variable 3	Variable 4
1	X	X	X	X
2	X	X	X	X
3	X	X	O	O
4	X	O	O	O
5	O	O	O	O

- Arbitrary pattern of missing data

The non-monotone (arbitrary) missing data pattern indicates situation when there is no particular pattern in the missing data structure.

Obs	Variable 1	Variable 2	Variable 3	Variable 4
1	X	X	X	O
2	O	X	O	X
3	X	O	X	X
4	X	X	X	O
5	O	X	X	X

.....and there are many other specific missing data patterns

For example,

- Missing by-design data pattern

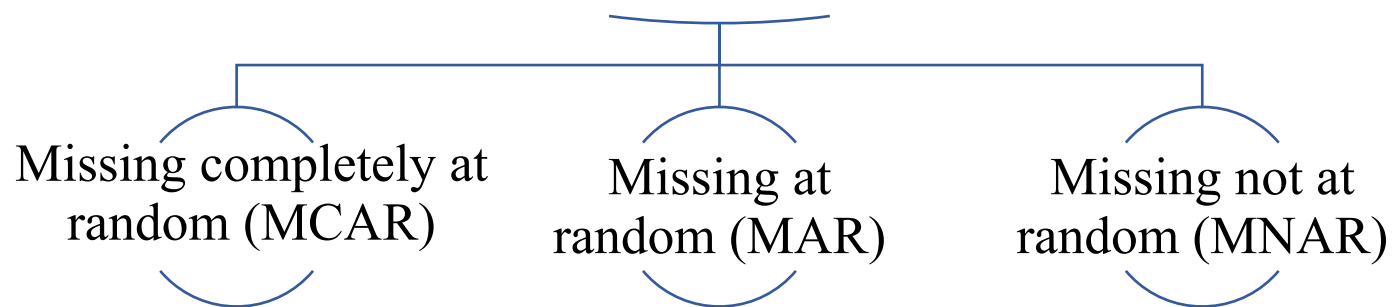
Obs	Variable 1	Variable 2	Variable 3	Variable 4
1	X	X	O	X
2	X	X	O	X
3	X	O	X	O
4	X	O	X	O
5	X	O	X	O

This type of missing data pattern arises in studies with randomization procedures to allow item missing data on selected variables for subsets of study observations



QUARTESIAN

Missing data mechanisms



Under MCAR the missing values are the simple random sample from all values, and there is not systematic rules of facing the missing data values.

Under MAR, the missing-data values do not contain any additional information given observed data about the missing-data mechanism, thus the process that causes missing data can be ignored.

Under MNAR there is the systematic difference after taking into account remains between observed values and missing values.



Strategies for dealing with missing data

Complete Case Analysis

- Involves list-wise removing of missing values.
- Is applicable when item missing data rates of 1% to 5% for single variables.

Complete Case Analysis with Weighting Adjustments

- This method with usage of weighting to compensate for missing data is generally limited to monotonic missing data patterns in which large numbers of variables are missing for each case.

Full Information Maximum Likelihood

- This method directly maximizes a likelihood function in which incomplete cases support only to estimation of parameters for which the sufficient statistics are functions of the observed values for the case.

Strategies of dealing with missing data

The expectation-maximization (EM) algorithm

- EM algorithm implements iterative two-step (expectation and maximization) approach to numerically derive ML estimates of parameters.

Single imputation

- Involves the usage of single imputed value replacing missing data on key variables.
- Established stochastic imputation method such as the hot deck, regression imputation, or predictive mean matching could be used for this.
- Median, mean, last or worst observations could be used to perform single imputation.

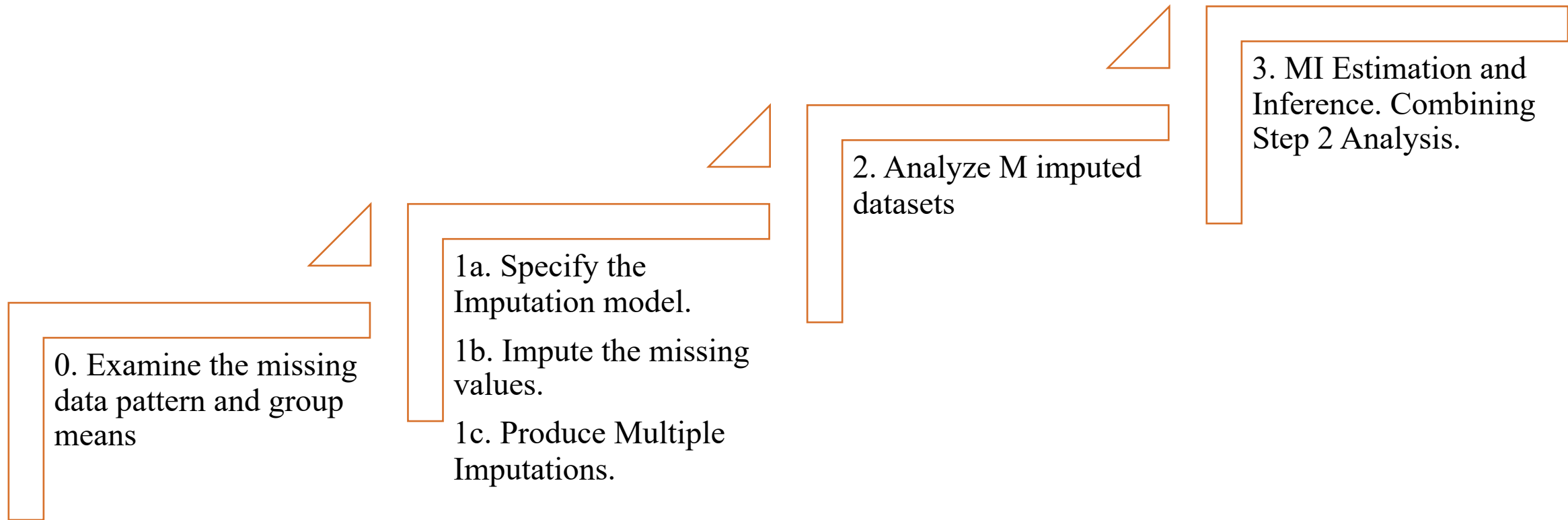
Multiple imputation

- Multiple imputation does not attempt to estimate each missing value through simulated values but rather to represent a random sample of the missing values.
- This process results in valid statistical inferences that properly reflect the uncertainty due to missing values; for example, valid confidence intervals for parameters.

Why multiple imputation?

- *MI is model-based.*
- *MI is stochastic.*
- *MI is multivariate.*
- *MI employs multiple independent repetitions* of the imputation procedure that permit the estimation of the uncertainty (the variance) in parameter estimates that is attributable to imputing missing values.
- *MI is robust* against minor departures from strict theoretical assumptions.
- *MI is very usable* in real statistical applications (SAS, STAT, R Studio, etc.).

Steps of performing multiple imputation



Imputation methods, dependending on missing data pattern and variable type

Missing Data Pattern	Variable Type	MI Method
Monotone	Continuous	Linear regression, predictive mean matching, propensity score
	Binary/ ordinary	Logistic regression
	Nominal	Discriminant function
Arbitrary	Continuous	With continuous covariates: MCMC monotone method MCMC full-data imputation
	Continuous	With mixed covariates: FCS regression* FCS predictive mean matching*
	Binary/ ordinary	FCS logistic regression*
	Nominal	FCS discriminant function*

* FCS MI involves specifying a series of univariate models for the conditional distribution of each partially observed variable given the other variables. This permits a great deal of flexibility, since an appropriate regression model can be selected for each variable (e.g. linear regression for continuous variables, logistic regression for binary variables).



Weekly P-NRS score will be set to missing after taking any rescue medication.

Analysis Endpoints

Primary Endpoint: Change from baseline in Pain Numeric Rating Scale at Week 5 (weekly value is calculated as average of all measures during the week), supposing that missing data is MAR.

Sensitivity Analysis I of Primary Endpoint: Change from baseline in Pain Numeric Rating Scale at Week 5, supposing the missings are MNAR (based on the reason for withdrawal).

Primary and sensitivity analysis will be performed using mixed effects model with repeated measures (MMRM). The model will contain treatment, week, and treatment-by-week interaction as fixed effects, baseline score and disease severity as covariates.

Step 0 (before MI) – testing missing pattern

It is very useful to examine the missing data pattern and group means using SAS PROC MI with NIMPUTE=0 and a SIMPLE option before imputing values.

Subject ID	TRTP	Severity	Baseline	Week 1	Week 2	Week 3	Week 4	Week 5
001	Placebo	Mild-Moderate	X	X	X	X	O	X
002	Placebo	Moderate-Severe	X	X	X	X	X	O
003	Active	Mild-Moderate	X	X	O	X	O	O
004	Placebo	Mild-Moderate	X	X	X	X	X	O
005	Active	Mild-Moderate	X	X	O	X	X	X
006	Active	Moderate-Severe	X	X	O	O	O	X
007	Placebo	Mild-Moderate	X	X	X	X	O	X



QUARTESIAN

Step 0 (before MI) – testing missing pattern

```
proc mi data=for_imputation nimpute=0 out= M_3 simple;  
var Week_1 - Week_5;  
run;
```

Missing Data Patterns

Group	Week_1	Week_2	Week_3	Week_4	Week_5	Freq	Percent	Group Means				
								Week_1	Week_2	Week_3	Week_4	Week_5
1	X	X	X	X	X	70	84.34	7.088946	6.232041	5.874014	5.719252	5.376531
2	X	X	X	X	.	3	3.61	7.476190	6.285714	6.515873	5.755556	
3	X	X	X	.	.	2	2.41	4.642857	4.404762	4.833333		
4	X	X	.	X	X	1	1.20	7.428571	8.600000		7.600000	6.666667
5	X	X	.	.	.	1	1.20	5.857143	5.000000			
6	X	.	X	X	X	1	1.20	8.400000		7.571429	7.000000	7.600000
7	X	.	.	.	.	4	4.82	7.271429				
8	O	O	O	O	O	1	1.20					

....so the missing data pattern is arbitrary

Step 1 - Multiple imputation: MAR

The Markov Chain Monte Carlo (MCMC) method will be used for implementing multiple imputation. This method is appropriate for non-monotonic missing data.

1. Getting the monotone missing pattern (Imputation of Intermittent missing data).

```
proc mi data=for_imputation nimpute=10 out= imput_1;  
    mcmc impute=monotone ;  
    var Severity Baseline Week_1 - Week_5;  
    by TRTP ;  
run;
```

Step 1 - Multiple imputation: MAR

2. Performing MI for monotone missing pattern data using regression

```
%sort(imput_1,_imputation_);  
  
proc mi data= imput_1 nimpute=1 out= imput_2;  
  by _imputation_;  
  class TRTP;  
  var TRTP Severity Baseline Week_1 - Week_5;  
  monotone reg;  
run;
```

Step 1 - Multiple imputation: MAR

3. Transposing data to get it vertically-oriented

```
%sort(imput_2, usubjid _imputation_);  
  
proc transpose data=imput_2 out=mar;  
    by usubjid _imputation_;  
    var baseline WEEK_1 WEEK_2 WEEK_3 WEEK_4 WEEK_5;  
run;
```

...The derivation type for this values will be defined as “MAR” in corresponding ADAM dataset



Step 1 - Multiple imputation: MNAR (based on reason for discontinuation)

MI with using regression assuming MNAR for subjects who discontinue from the study due to usage of any rescue medication or due to adverse event. And MI assuming MAR for subjects without such conditions will be performed.

Missing Pain Numeric Rating Scale results for subjects who withdraw due to AE or who took any rescue drug will be imputed based on baseline distribution of weekly scores for all subjects with the same disease severity status assuming trimmed normal (from 5 to 10).

Step 1 - Multiple imputation: MNAR (based on reason for discontinuation)

1. Selecting only the data for patients who did not withdraw early due to an AE or taking rescue medication.

```
%sort(imput_1 imput_2, usubjid);  
  
data not_ae_rm ;  
    merge imput_2(in=a) adsl(in=b where=(lowercase(DSCREAS)='adverse event' or  
RESUSFL='Y') keep=usubjid DSCREAS RESUSFL);  
    by usubjid;  
    if a=1 and b=0 then output not_ae_rm ;  
run;
```

Step 1 - Multiple imputation: MNAR

2. Selecting records for subjects who withdrew early due to an AE or took opioid rescue medication. - MNAR

```
data ae_rm ;  
    merge imput_1(in=a) adsl(in=b where=(lowercase(DSCREAS)='adverse event' or RESUSFL ='Y'))  
keep=usubjid DCSREAS RESUSFL);  
    by usubjid;  
    if a=1 and b=1;  
run;
```

3. Combining two datasets

```
data all_for_mnar;  
    set ae_rm non_ae_rm;  
run;
```

Step 1 - Multiple imputation: MNAR

4. Transposing data to get it vertically-oriented and specifying baseline values to be used in proc MI

```
%sort(all_for_mnar,usubjid _imputation_ Severity TPTP);
```

```
proc transpose data= all_for_mnar out=ae_mnar ;  
    by usubjid _imputation_ Severity TPTP;  
    var Baseline-- Week_5;  
run;
```

```
data base_ae_rm_mnar;  
    set ae_mnar;  
    if _NAME="Baseline" then new_base=coll;  
run;
```

Step 1 - Multiple imputation: MNAR

5. Performing MI with using monotone regression to impute missings with values from a truncated normal distribution 5 to 10.

```
%sort(base_ae_rm_mnar , _imputation_);  
  
proc mi data=base_ae_rm_mnar out=mnar_ nimpute=1 min=. 5 max=. 10;  
  by _imputation_;  
  class ADDISSEVN;  
  var ADDISSEVN new_base;  
  monotone reg(new_base=ADDISSEVN);  
run;
```

...The derivation type for this values will be defined as “MNAR” in corresponding ADAM dataset

Step 2 - Mixed effects model with repeated measures (MMRM)

This model contains treatment, visit, and treatment-by-visit interaction as fixed effects, baseline score and disease severity as covariates

The appropriate derivation type (MAR or MNAR) is used separately to get the results from the following model

```
proc mixed data = data_for_analysis;  
  by _imputation_ ;  
  class trtp visit usubjid severity;  
  model chg = trtp visit trtp*avisit baseline Severity /ddfm=kr;  
  repeated visit/ type = un subject = usubjid;  
  lsmeans trtp*visit / pdiff cl;  
  
run;
```



Step 3 – Combing the results, using PROC MIANALYZE

```
Proc mianalyze data=ls;  
  by visit trtp;  
  modeleffects estimate;  
  stderr stderr;  
Run;
```

Proc mianalyze is applied for LS Means as well as for LS Mean differences

```
proc mianalyze data=diff;  
  by avisitn trtp _trtp;  
  modeleffects estimate;  
  stderr stderr;  
run;
```

Results of the Primary and Sensitivity Analyses

	Results under MAR		Results under MNAR	
	Treatment A	Treatment B	Treatment A	Treatment B
LS Mean (SE)	-1.995 (0.327)	-2.452 (0.310)	-1.958 (0.329)	-2.145 (0.310)
95 % CI	(-2.636, -1.354)	(-3.061, -1.843)	(-2.603, -1.312)	(-2.754, -1.536)
LS Mean difference A vs B (SE)	0.456 (0.421)	x	0.187 (0.426)	x
95 % CI	(-0.369, 1.283)	x	(-0.647, 1.022)	x

Conclusion

- MI is a very useful tool for dealing with missing data.
- This method gives the robust parameter estimates.
- It is applicable when other strategies for dealing with missing data could not be applied.
- It uses different methods, depending on the type of missing variable and pattern.
- It allows implementing sensitivity analysis of primary endpoints, assuming another missing data mechanism or using another imputation method.



Thank you!

Daria Shamaida
daria.shamaida@quartesian.com
Kyiv, Ukraine