IMS Health & Quintiles are now IQVIA®

Multiple Imputation Methods for Handling Missingness

28th March 2019

Jack Elkes and Sophie Hodge

+ What is MI?
+ Assumptions for MI
+ The different missing mechanisms
+ Variables used to impute missingness - what assumptions are made
+ LOCF vs MI
+ Patterns of Missingness
+ Implementation in SAS

What is MI?

Multiple imputation is a method to handle missing data. It allows estimation across more observations and therefore reduces biased towards observed results.

- Multiple imputation for missing data makes it possible to obtain approximately asymptotically unbiased estimates of all the parameters from the random error.

- This multiple imputation for missing data allows us to obtain good estimates of the standard errors. The multiple imputation for missing data is unlike single imputation, since it doesn’t allow additional error to be introduced.

- We can perform multiple imputation for missing data with any kind of data in any kind of analysis, without well-equipped software.

What is MI? (contd.)

Multiple imputation is a method to handle missing data. It allows estimation across more observations and therefore reduces biased towards observed results.

In multiple imputation we:

1. Create a number, M, of imputed datasets
2. Estimate our parameter(s) of interest within each imputed dataset
3. Aggregate the results of these M analyses

Estimating variances and finding confidence intervals is relatively easy if we create multiple imputations, but is rather difficult with only a single imputation.

The different missing mechanisms

Missing completely at random (MCAR)

- Whether a subject has missing data is completely unrelated to the other information in the data.
  - This is extremely rare but also means the analysis performed provide unbiased results.
  - E.g. Missing scan data due to scan being lost at the site

Missing at random (MAR)

- Other variables (but not the variable which is missing itself) in the dataset can be used to predict missingness.
  - E.g. Missing scan data due to subject missing a visit due to an Adverse Event.

Missing not at random (MNAR)

- The unobserved value of the variable with missing predicts missingness.
  - E.g. Missing scan data due to subject dropping out due to lack of efficacy
  - This means the data will not help in explaining how those with missing data are different, therefore we would need to perform a sensitivity analysis to explore and compare the outcomes under different assumptions.

Variables used to impute missingness - what assumptions are made

- Usually, all variables which will be used in our model of interest / analysis model should be included in the imputation model.

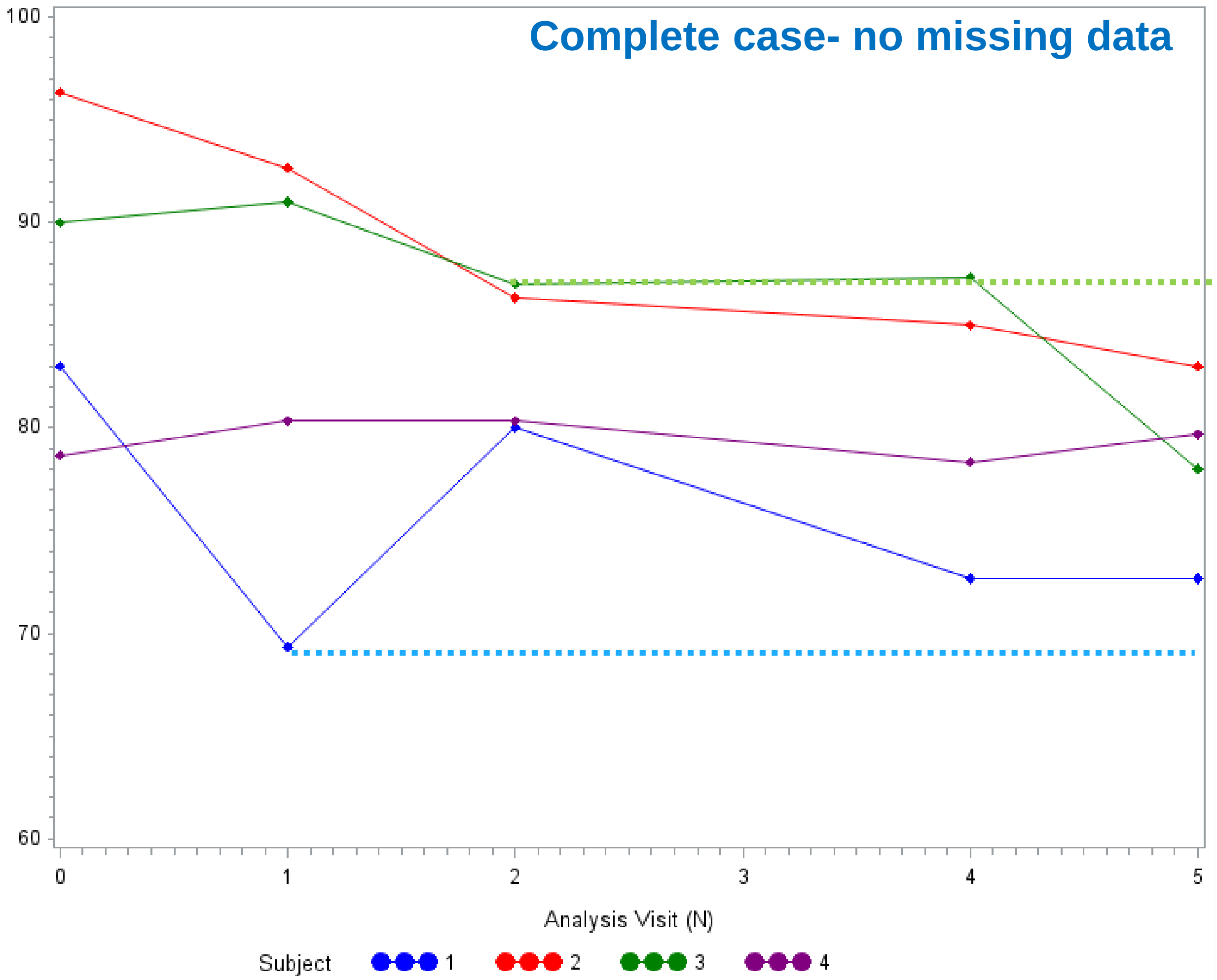
- If we are imputing missing covariates, the outcome variable **must** be included, to ensure that the imputed covariate values have the correct association with the outcome.

LOCF vs MI

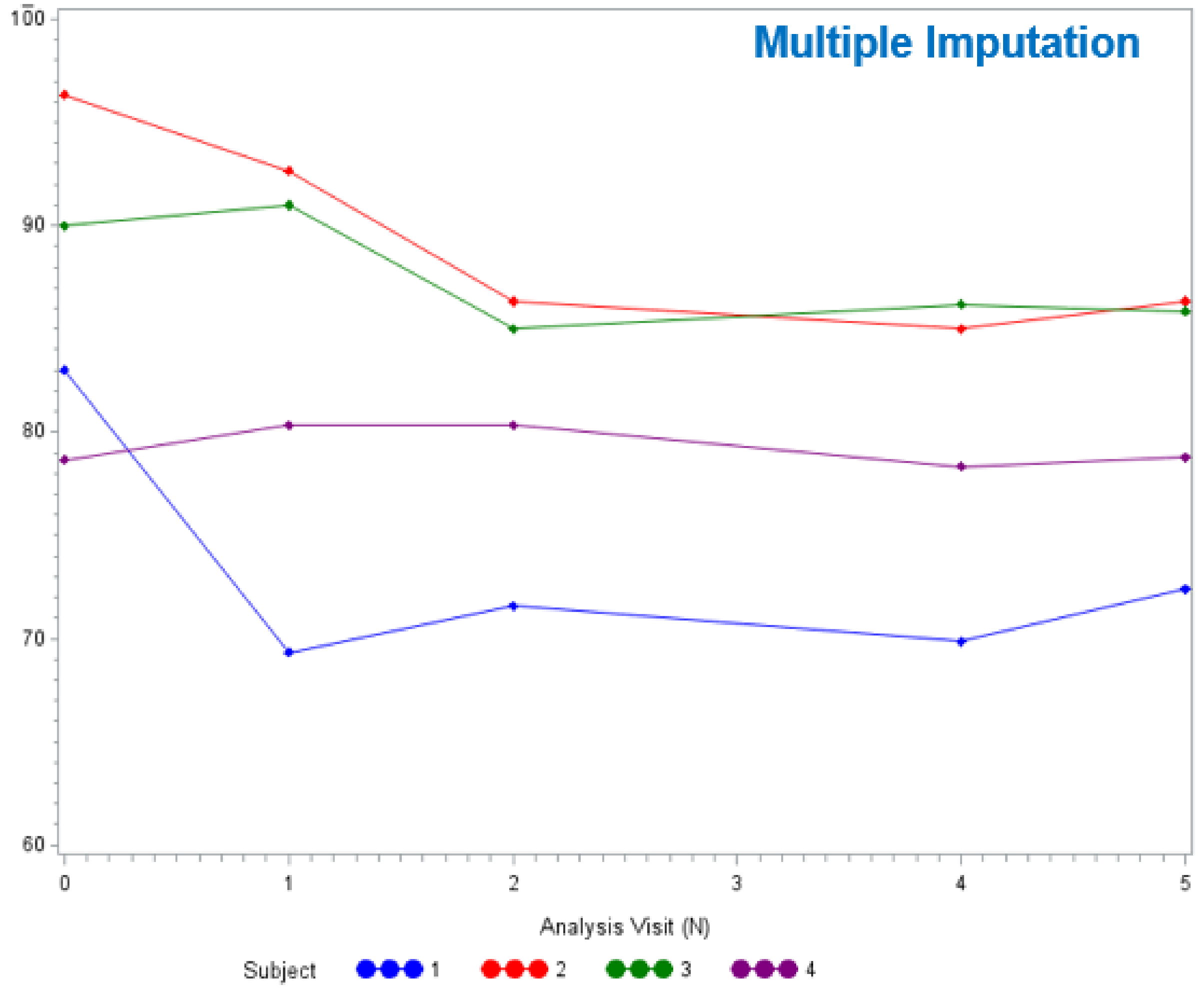
- Last observation carried forward (LOCF)- last observation before a subject discontinues is likely to represent when a patient is at a low point in terms of their heath and can therefore be considered a conservative estimate of treatment efficacy.

- Does not use information from other subjects in the study.

Example



Complete case- no missing data
100
90
80
70
60
0
1
2
3
4
5
Analysis Visit (N)
Subject 1 2 3 4



Multiple Imputation
100
90
80
70
60
0
1
2
3
4
5
Analysis Visit (N)
Subject 1 2 3 4

Patterns of Missingness

Monotone

usubjid	TRT02AN	score_0	score_1	score_2	score_4	score_5
-102-02	2	80	72	74.666666667	76.666666667	64.33333333
-102-05	5	87	66.666666667	67	64.666666667	69
-103-02	1	75.333333333	78.666666667	66.333333333	74.333333333	.
-112-01	2	90	91	.	.	.
-112-02	3	94.333333333	80.333333333	86.666666667	.	.
-112-06	3	96.333333333	82.333333333	88	88.666666667	77
-112-12	1	87	84	.	.	.
-112-13	5	96.666666667	75.333333333	79.333333333	78.333333333	79.33333333
-115-01	2	95.666666667	83.333333333	68	64.666666667	85.33333333

Patterns of Missingness

Non-Monotone

usubjid	TRT02AN	score_0	score_1	score_2	score_4	score_5
-115-05	2	91.333333333	81	75.666666667	80.333333333	79
-120-01	5	85.666666667	65.666666667	.	72	61
-121-05	5	90.333333333	74.666666667	.	.	65
-121-07	1	80.666666667	67	59.666666667	57.333333333	71
-124-04	3	92.666666667	77.666666667	.	.	75.666666667
-130-01	3	91	79.333333333	80.666666667	76.333333333	91.666666667
-130-03	1	88.666666667	91.333333333	.	89	83
-130-04	5	88	87	.	.	84
-130-07	1	90.333333333	89.666666667	.	.	84

Example study

- Phase 3 Study Hypertension study in paediatric patients comparing 3 dosing levels with placebo
- Primary endpoint: change in diastolic blood pressure (DBP) from baseline to visit 5
- Age group (<5, >=5), gender (M, F) are used as covariates in the primary analysis

usubjid	score_0	score_1	score_2	score_4	score_5
-102-02	80	72	74.666666667	76.666666667	64.333333333
-102-05	87	66.666666667	67	64.666666667	69
-103-02	75.333333333	78.666666667		74.333333333	71.333333333
-112-01	90	91			78
-112-02	94.333333333	80.333333333			
-112-06	96.333333333	82.333333333	88	88.666666667	77
-112-12	87	84			81
-112-13	96.666666667	75.333333333	79.333333333	78.333333333	79.333333333
-115-01	95.666666667	83.333333333	68	64.666666667	85.333333333
-115-05	91.333333333	81	75.666666667	80.333333333	79

Implementation in SAS

Step 1: PROC MI will be applied to create monotone data.

SUBJID – subject ID number
TRT1 – Treatment Arm [Placebo]
TRT2 – Treatment Arm [Active treatment x]
TRT3 – Treatment Arm [Active treatment y]
TRT4 – Treatment Arm [Active treatment z]
SCORE_0 - AVAL at baseline
SCORE_1 - post-baseline visit 1
SCORE_2 - post-baseline visit 2
SCORE_4 –post-baseline visit 4
SCORE_4 –post-baseline visit 5
Sexn- Gender (0=male, 1=female)
Agegr- Age Group (0= <50, 1= >=50)

usubjid	TRT02AN	score_0	score_1	score_2	score_4	score_5
-102-02	2	80	72	74.666666667	76.666666667	64.333333333
-102-05	5	87	66.666666667	67	64.666666667	69
-103-02	1	75.333333333	78.666666667	69.028480903	74.333333333	71.333333333
-112-01	2	90	91	82.451260903	84.874511117	78
-112-02	3	94.333333333	80.333333333			
-112-06	3	96.333333333	82.333333333	88	88.666666667	77
-112-12	1	87	84	89.204518302	95.219750642	81
-112-13	5	96.666666667	75.333333333	79.333333333	78.333333333	79.333333333
-115-01	2	95.666666667	83.333333333	68	64.666666667	85.333333333
-115-05	2	91.333333333	81	75.666666667	80.333333333	79

```
proc mi data=datain out=datain_mono nimpute=100 seed=123;
     var trt1 trt2 trt3 agegr sexn score_0 score_1 score_2 score_4 score_5 ;
     mcmc chain=multiple impute=monotone ;
run;
```

Implementation in SAS

Step 2: PROC MI will be applied to the monotone data to complete the imputation.

```
proc mi data=datain_mono out=datain_reg seed=465 nimpute=1;
  by _imputation_ trt;
 var trt1 trt2 trt3 agegr sexn score_0 score_1 score_2 score_4 score_5;
 class trt1 trt2 trt3 agegr sexn;
 monotone regression;
run;
```

usubjid	score_0	score_1	score_2	score_4	score_5
-102-02	80	72	74.666666667	76.666666667	64.333333333
-102-05	87	66.666666667	67	64.666666667	69
-103-02	75.333333333	78.666666667	69.028480903	74.333333333	71.333333333
-112-01	90	91	82.451260903	84.874511117	78
-112-02	94.333333333	80.333333333	103.61477045	92.566662527	86.619715927
-112-06	96.333333333	82.333333333	88	88.666666667	77
-112-12	87	84	89.204518302	95.219750642	81
-112-13	96.666666667	75.333333333	79.333333333	78.333333333	79.333333333
-115-01	95.666666667	83.333333333	68	64.666666667	85.333333333
-115-05	91.333333333	81	75.666666667	80.333333333	79

Implementation in SAS

Step 3: Step 3: Transform the imputed dataset so that for each subject, we will compute a change from baseline to visit 5 for DBP

```
data datain_reg1;
  set datain_reg;
  timeptn=1;
  score_c=score_5 - score_0;
run;
```

usubjid	score_0	score_1	score_2	score_4	score_5	score_c
102-02	80	72	74.666666667	76.666666667	64.333333333	-15.66666667
102-05	87	66.666666667	67	64.666666667	69	-18
103-02	75.333333333	78.666666667	69.028480903	74.333333333	71.333333333	-4
112-01	90	91	82.451260903	84.874511117	78	-12
112-02	94.333333333	80.333333333	103.61477045	92.566662527	86.619715927	-7.713617406
112-06	96.333333333	82.333333333	88	88.666666667	77	-19.33333333
112-12	87	84	89.204518302	95.219750642	81	-6
112-13	96.666666667	75.333333333	79.333333333	78.333333333	79.333333333	-17.33333333
115-01	95.666666667	83.333333333	68	64.666666667	85.333333333	-10.33333333
115-05	91.333333333	81	75.666666667	80.333333333	79	-12.33333333

Implementation in SAS

Step 4: PROC MIXED

```
proc mixed data=datain_reg1;
  by _imputation_ timeptn;
  class trt02pn agegr sexn;
  model score_c = trt02pn agegr sexn score_0/solution covb;
  lsmeans trt02pn / pdiff diff alpha= 0.05 cl;

  ods output lsmeans=est_mi;
Run;
```

Imputation	trt01pn	Estimate	StdErr	Lower	Upper
1	1	-10.6503	1.5127	-13.6418	-7.6587
1	2	-10.3218	1.3598	-13.0110	-7.6326
1	3	-11.4794	1.4133	-14.2742	-8.6845
1	5	-12.3369	1.6210	-15.5425	-9.1313
2	1	-10.2247	1.4778	-13.1471	-7.3023
2	2	-10.5502	1.3284	-13.1771	-7.9232
2	3	-11.1236	1.3806	-13.8538	-8.3934
2	5	-12.2189	1.5835	-15.3503	-9.0874
3	1	-11.2351	1.5105	-14.2223	-8.2479
3	2	-10.7789	1.3579	-13.4642	-8.0937

Implementation in SAS

Step 5: Results from all ANCOVA analysis are then combined together for overall inference using PROC MIANALYZE

```
proc mianalyze data=est_mi;
   ods output ParameterEstimates=parms;
   modeleffects estimate;
   stderr Stderr;
run;
```

Estimate	StdErr	LCLMean	UCLMean
-11.048197	1.731946	-14.4436	-7.65284

References

- https://stats.idre.ucla.edu/sas/seminars/multiple-imputation-in-sas/mi_new_1/
- http://stats.idre.ucla.edu/sas/seminars/multiple-imputation-in-sas/mi_new_1/
- http://stats.idre.ucla.edu/wp-content/uploads/2017/01/Missing-Data-Techniques_UCLA.pdf
- http://thestatsgeek.com/2013/07/06/when-is-complete-case-analysis-unbiased/
- https://stats.idre.ucla.edu/sas/seminars/multiple-imputation-in-sas/mi_new_1/
- https://www.statisticssolutions.com/multiple-imputation-for-missing-data/
- https://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_mi_sect021.htm
- https://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_mianalyze_sect012.htm
- O'Kelly, M. and Ratitch, B. (2014). *Clinical trials with missing data*. 1st ed. John Wiley & Sons, Ltd.

IMS Health & Quintiles are now IQVIA®

Any Questions?