
BEYOND MISSING DATA PATTERNS: AUGMENTING PROC MI TABLES WITH SUMMARY OF IMPUTED DATA

Lamine Signate, Covance, Leiden, The Netherlands

Domenico Magazzù, Covance, Milan, Italy

Tommi Nurminen, Tale Ltd, Jyväskylä, Finland

PHUSE EU Connect 2019, 12 Nov 2019, Amsterdam

Agenda

- ▶ Introduction to SAS PROC MI (Multiple Imputation) and Data Patterns
- ▶ Objective
- ▶ Algorithm and Programming Steps
 - Case example
 - Augmenting PROC MI Table with Summary of Imputed Data
- ▶ Results
 - Presentation of the PROC MI Table with Summary Statistics of Imputed Data
- ▶ Conclusion

Introduction to SAS PROC MI and Data Patterns (1/3)

- ▶ Longitudinal studies give rise to missing data due to dropouts and a range of other reasons.
 - Some observed data may be censored due to e.g. use of rescue therapy or treatment switches
- ▶ To overcome the statistical issues inherent to the amount of missing data, appropriate statistical analysis and inference may require that missing data are imputed using some imputation model
- ▶ In multiple imputation (MI), each missing value is replaced with a set of plausible values that reflect the uncertainty about the true value (*Little & Rubin, 2002*)

Introduction to SAS PROC MI and Data Patterns (2/3)

► “The SAS MI procedure performs multiple imputation of missing data” (SAS institute Inc.)

- PROC MI can be used for summarising the frequencies of each unique missing data pattern (MDP) in the dataset along with the mean values of the analysis variables within each pattern

Missing Data Patterns						Group Means		
Group	Oxygen	RunTime	RunPulse	Freq	Percent	Oxygen	RunTime	RunPulse
1	X	X	X	21	67.74	46.353810	10.809524	171.666667
2	X	X	.	4	12.90	47.109500	10.137500	.
3	X	.	.	3	9.68	52.461667	.	.
4	.	X	X	1	3.23	.	11.950000	176.000000
5	.	X	.	2	6.45	.	9.885000	.

```
proc mi data=FitMiss nimpute=0 out=outmi;  
var Oxygen RunTime RunPulse;  
run;
```

Table extracted from SAS support:

https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_mi_sect002.htm

Accessed on: 15 Oct 2019

Introduction to SAS PROC MI and Data Patterns (3/3)

- ▶ PROC MI displays only summary statistics of available data
 - However, missing data from studies are often imputed depending on the study objectives

Missing Data Patterns						Group Means		
Group	Oxygen	RunTime	RunPulse	Freq	Percent	Oxygen	RunTime	RunPulse
1	X	X	X	21	67.74	46.353810	10.809524	171.666667
2	X	X	.	4	12.90	47.109500	10.137500	.
3	X	.	.	3	9.68	52.461667	.	.
4	.	X	X	1	3.23	.	11.950000	176.000000
5	.	X	.	2	6.45	.	9.885000	.

```
proc mi data=FitMiss nimpute=0 out=outmi;  
var Oxygen RunTime RunPulse;  
run;
```

Table extracted from SAS support:

https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_mi_sect002.htm

Accessed on: 15 Oct 2019

Objective

The objective of our work was to augment PROC MI tables with summary of imputed data

Missing Data Patterns						Group Means		
Group	Oxygen	RunTime	RunPulse	Freq	Percent	Oxygen	RunTime	RunPulse
1	X	X	X	21	67.74	46.353810	10.809524	171.666667
2	X	X	X _{imp}	4	12.90	47.109500	10.137500	A _{imp}
3	X	X _{imp}	X _{imp}	3	9.68	52.461667	B _{imp}	C _{imp}
4	X _{imp}	X	X	1	3.23	D _{imp}	11.950000	176.000000
5	X _{imp}	X	X _{imp}	2	6.45	E _{imp}	9.885000	F _{imp}

Table extracted from SAS support:

https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_mi_sect002.htm. Accessed on: 15 Oct 2019

Algorithm and Programming Steps

Case example (1/2)

- ▶ **Data Source:** Non-interventional, single-arm study assessing the effectiveness of a specific pharmaceutical intervention in patients with axial spondyloarthritis (axSpA) in daily practice
 - Primary endpoint was the Change from Baseline to Week 52 in Bath Ankylosing Spondylitis Disease Activity Index (BASDAI)
 - BASDAI was assessed at 4 time points: Baseline, week 12, week 24 and week 52
 - For the purposes of the current work, data were assumed to be missing at random (MAR)
 - The MI was done in 2 parts: intermittent missing data imputed using Markov chain Monte Carlo (MCMC) and monotone missing data imputed using regression model
 - Further details of the imputation model are not reported as out of the scope for the current work

Algorithm and Programming Steps

Case example (2/2)

```
ods output misspattern=missp;
proc mi nimpute=0 data=miss_data;
var week0 week12 week24 week52;
run;
```

Table 1. PROC MI output using initial data

Pattern	Subjects with Missing Data Pattern n (%)	Observed Missing Data pattern				BASDAI Mean				
		Baseline	Week 12	Week 24	Week 52	Baseline	Week 12	Week 24	Week 52	
1	413 (73.2)	X	X	X	X	6.0	3.5	3.3	3.1	Complete data
2	62 (11.0)	X	X	X	O	6.5	5.1	5.5	NC	Monotone MDP
3	60 (10.6)	X	X	O	O	6.1	5.1	NC	NC	
4	17 (3.0)	X	X	.	X	6.7	4.1	NC	4.4	Intermittent MDP
5	7 (1.2)	X	.	X	X	6.2	NC	4.0	3.6	Intermittent + Monotone MDP
6	3 (0.5)	X	.	X	O	5.5	NC	5.7	NC	
7	2 (0.4)	X	.	.	X	8.8	NC	NC	4.6	

Algorithm and Programming Steps

Augmenting PROC MI Table with Summary of Imputed Data (1/6)

- **Step 1:** To derive MDP manually using a SAS data step
 - Output dataset 'miss_pattern' contains all the subjects and their associated MDP

```
data miss_pattern;
    set miss_data;
    array missx (*) $ week0_miss week12_miss week24_miss week52_miss;
    array obsx (*) $ week0 week12 week24 week52;
    do i=1 to dim(missx);
        if missing(obsx[i]) then missx[i]='M'; /* missing values are noted by M */
        else missx[i]='X';
    end;
    /* MDP derivation */
    miss_pat=cats (week0_miss, week12_miss, week24_miss, week52_miss);
run;
```

Algorithm and Programming Steps

Augmenting PROC MI Table with Summary of Imputed Data (2/6)

- ▶ **Step 2:** To perform MI using the input dataset 'miss_pattern' generated in Step 1
 - Use PROC MI to conduct MI
 - Output dataset 'miout1' contains complete data and all types of MDP (variable=*miss_pat*)
 - Figure below shows a sample of the different imputed values for a same subject
- i Any imputation method including single-value imputation could be used instead of MI in this step

Imputation Number	WEEK0	WEEK12	WEEK24	WEEK52	WEEK0_MISS	WEEK12_MISS	WEEK24_MISS	WEEK52_MISS	MISS_PAT
1	5.8	2.4	1.2534	2.0428	X	X	M	M	XXMM
2	5.8	2.4	1.5803	3.1348	X	X	M	M	XXMM
3	5.8	2.4	5.0146	1.3532	X	X	M	M	XXMM
4	5.8	2.4	2.2591	3.3814	X	X	M	M	XXMM
5	5.8	2.4	5.2887	3.4578	X	X	M	M	XXMM
6	5.8	2.4	2.8209	2.9045	X	X	M	M	XXMM
7	5.8	2.4	3.3606	2.9605	X	X	M	M	XXMM
8	5.8	2.4	1.6949	1.721	X	X	M	M	XXMM
9	5.8	2.4	0.7946	0.0698	X	X	M	M	XXMM
10	5.8	2.4	0.9307	1.0539	X	X	M	M	XXMM

Fig. 'miout1' dataset: a sample including MI for a subject with missing values at Week 24 and 52.

Algorithm and Programming Steps

Augmenting PROC MI Table with Summary of Imputed Data (3/6)

- ▶ **Step 3:** To compute the mean values of complete data after MI by type of MDP
 - Use of input dataset 'miout1' and
 - Create a dataset 'means' containing the mean value at each time point for each type of MDP
 - The objective was to obtain the same means for observed values as in Table 1 and with NC values replaced with means of imputed values

```
proc sql;  
  create table means  
  as select      miss_pat,  
                 mean(week0) as mean_week0,  
                 mean(week12) as mean_week12,  
                 mean(week24) as mean_week24,  
                 mean(week52) as mean_week52  
  from miout1  
  group by miss_pat;  
quit;
```

Algorithm and Programming Steps

Augmenting PROC MI Table with Summary of Imputed Data (4/6)

- ▶ **Step 4:** To compute the frequencies (n) and percent (%) by type of MDP using PROC SQL
 - Create a dataset 'freq' containing the n and % that were the same as those generated by PROC MI
- i This step could also be done by an SQL step directly from 'miout1' on imputation number using variable `_imputation_ =1` and using the 'distinct' clause

```
proc sql;
  create table cnt1
  as select miss_pat, count(distinct usubjid) as count, joinkey=1
  from miout1 where _imputation_=1 group by miss_pat;

  create table bign
  as select count (distinct usubjid) as bign, joinkey=1
  from miout1 where _imputation_=1;

  create table freq as select a.*, b.bign ,count/bign *100 as percent
  from cnt1 a
  left join bign b on a.joinkey=b.joinkey;
quit;
```

Algorithm and Programming Steps

Augmenting PROC MI Table with Summary of Imputed Data (5/6)

- ▶ **Step 5:** To create a shell keeping one occurrence of each MDP
 - Shell also contained the indicators of whether each analysis variable was missing or observed for the particular MDP

```
proc sort nodupkey data=miout1  
out=shell (keep= week0_miss week12_miss week24_miss week52_miss);  
    by miss_pat;  
run;
```

Algorithm and Programming Steps

Augmenting PROC MI Table with Summary of Imputed Data (6/6)

- ▶ **Step 6:** To merge all datasets (i.e. 'freq', 'means', and 'shell') including n, % and means of observed and imputed data by each MDP
 - Means of imputed data were annotated with square brackets so numeric means needed to be rounded and converted to characters
 - i The additional SAS code to distinguish between intermittent and monotone missing data is not shown

```
data final;
    merge means freq shell;
    by miss_pat;
    length weekc0 weekc12 weekc24 weekc52 $30;
    array missx (*) $ week0_miss week12_miss week24_miss week52_miss;
    array visx (*) mean_week0 mean_week12 mean_week24 mean_week52;
    array vischarx (*) $ weekc0 weekc12 weekc24 weekc52;
    do i=1 to dim(missx);
        if missx[i] ne 'M' then vischarx[i]=put(round(visx[i],0.1),4.1);
        else vischarx[i]='[' || put(round(visx[i],0.1),4.1) || ']';
    end;
run;
```

Results: *Presentation of the PROC MI Table with Summary Statistics of Imputed Data*

- The final dataset obtained in Step 6 was displayed using a PROC REPORT

Table 2: Missing data patterns with means of observed and imputed data

		Observed Missing Data Pattern				BASDAI Mean over Imputed Data				
Pattern	Subjects with Missing Data		Baseline	Week 12	Week 24	Week 52	Baseline	Week 12	Week 24	Week 52
	Pattern	n (%)								
1	413	(73.2)	X	X	X	X	6.0	3.5	3.3	3.1
2	62	(11.0)	X	X	X	O	6.5	5.1	5.5	[4.7]
3	60	(10.6)	X	X	O	O	6.1	5.1	[4.3]	[4.1]
4	17	(3.0)	X	X	.	X	6.7	4.1	[4.2]	4.4
5	7	(1.2)	X	.	X	X	6.2	[4.1]	4.0	3.6
6	3	(0.5)	X	.	X	O	5.5	[4.8]	5.7	[5.0]
7	2	(0.4)	X	.	.	X	8.8	[5.6]	[4.9]	4.6

Conclusion

- ▶ Sequence of programming steps to augment the MDP summary from PROC MI with summaries of imputed data was presented
 - Some of the programming steps could be eliminated if the pattern number related to each study subject could be retrieved directly from PROC MI
- ▶ The final output (Table 2) makes it easy to compare the behaviour of model-based imputations with some old-school methods such as LOCF
- ▶ Presenting a summary including the means for the observed data and the imputed data by MDP may help to evaluate the appropriateness of imputed data and thus the overall conclusions

Questions



References

- Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (Second ed.). London: Wiley
- European Medicines Agency (2010). Guideline on Missing Data in Confirmatory Clinical Trials.
- SAS Institute Inc. (2011). SAS/STAT® 9.3 User's Guide. Cary, NC: SAS Institute Inc.

Acknowledgment

We thank Covance for their funding in publishing this work

We thank Frederike de Vries for her support and guidance for this work

We thank all the audience today!