

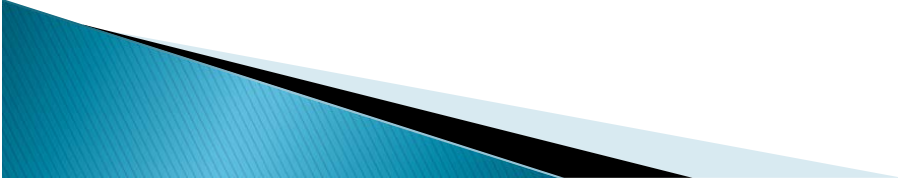


PhUSE
Annual Conference 2014
Paper SP04



Multiple Imputation: Better than Single Imputation in Pain Studies?

Author: Ashik Chowdhury
CYTEL, Pune, India



London
12th – 15th October 2014

Disclaimer

Any comments or statements made herein solely those of the author and do not necessarily reflect the views of the company.

Presentation Plan

Part I: Missing data

- Why missing data matters?
- Key points in regulatory guidance
- Scenario in industry

Part II: Multiple Imputation Technique

- Steps
- SAS PROCs
- Methods available in PROC MI

Part III: Comparison of methods

- Based on a simulation study

Presentation Plan

Part I: Missing data

- Why missing data matters?
- Key points in regulatory guidance
- Scenario in industry

Part II: Multiple Imputation Technique

- Steps
- SAS PROCS
- Methods available in PROC MI

Part III: Comparison of methods

- Based on a simulation study

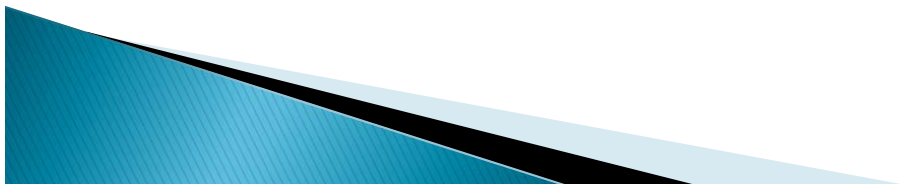
Part I: Missing Data

Why does it matter?

- ☐ May introduce bias in results
- ☐ Reduces power

Ideal situation: Achieve complete data

Possible solution: Imputation



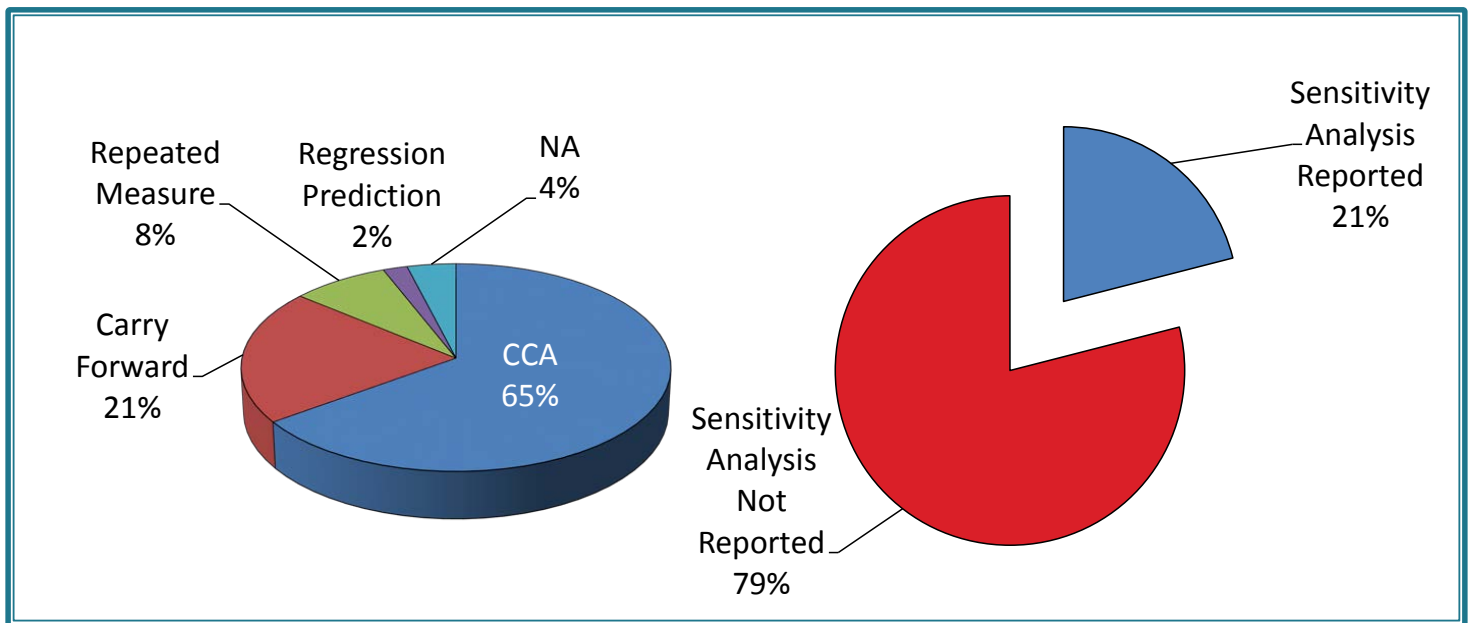
Part I: Missing Data

Key points in regulatory guidance about Imputation

- ☐ No single correct method to handle
- ☐ Describe in advance how missing data will be handled
- ☐ Sensitivity analysis
- ☐ LOCF and BOCF should not be used as the primary approach unless the assumptions that underlie them are scientifically justified

Part I: Missing Data

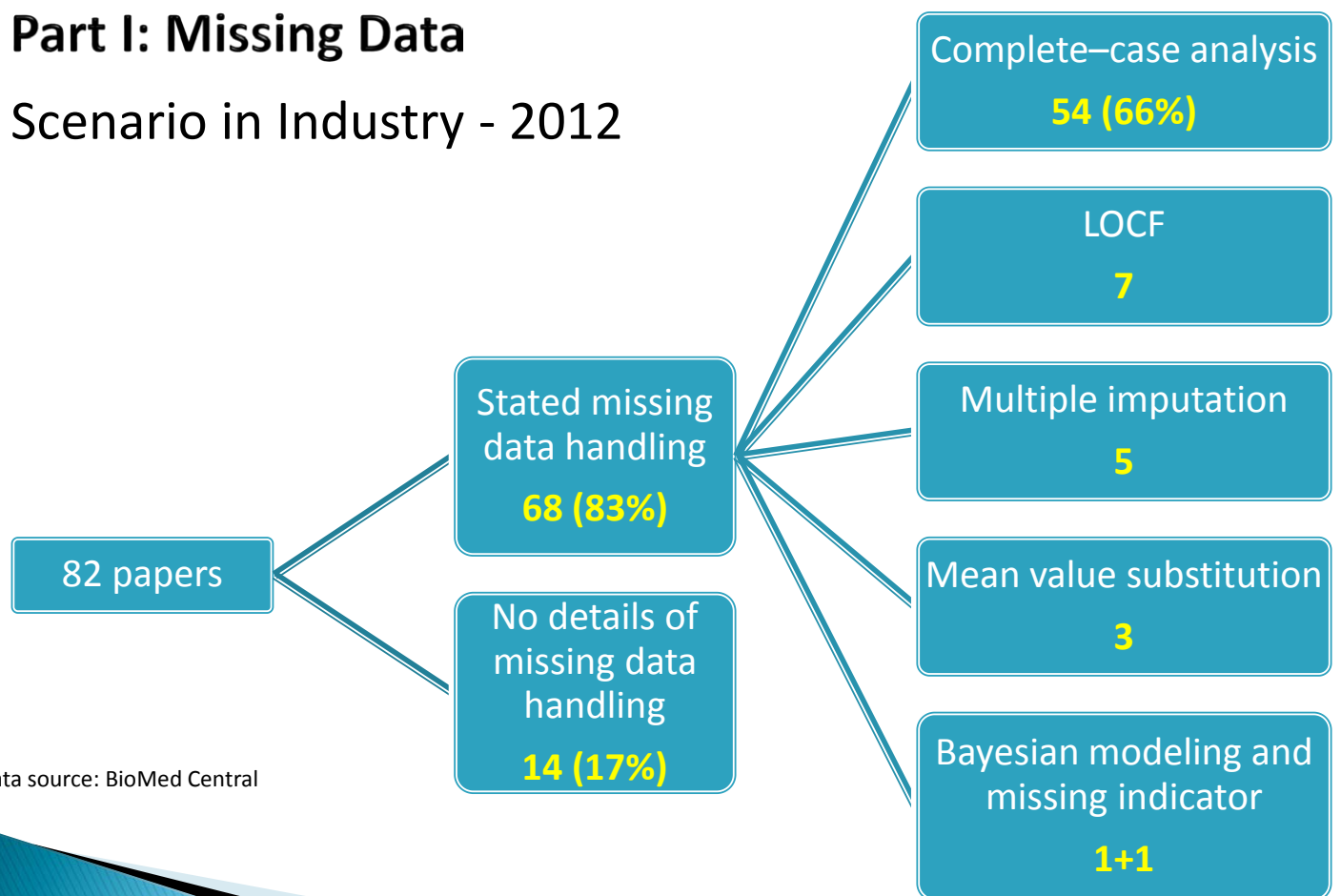
Scenario in Industry - 2004



Data source: SAGE Publication

Part I: Missing Data

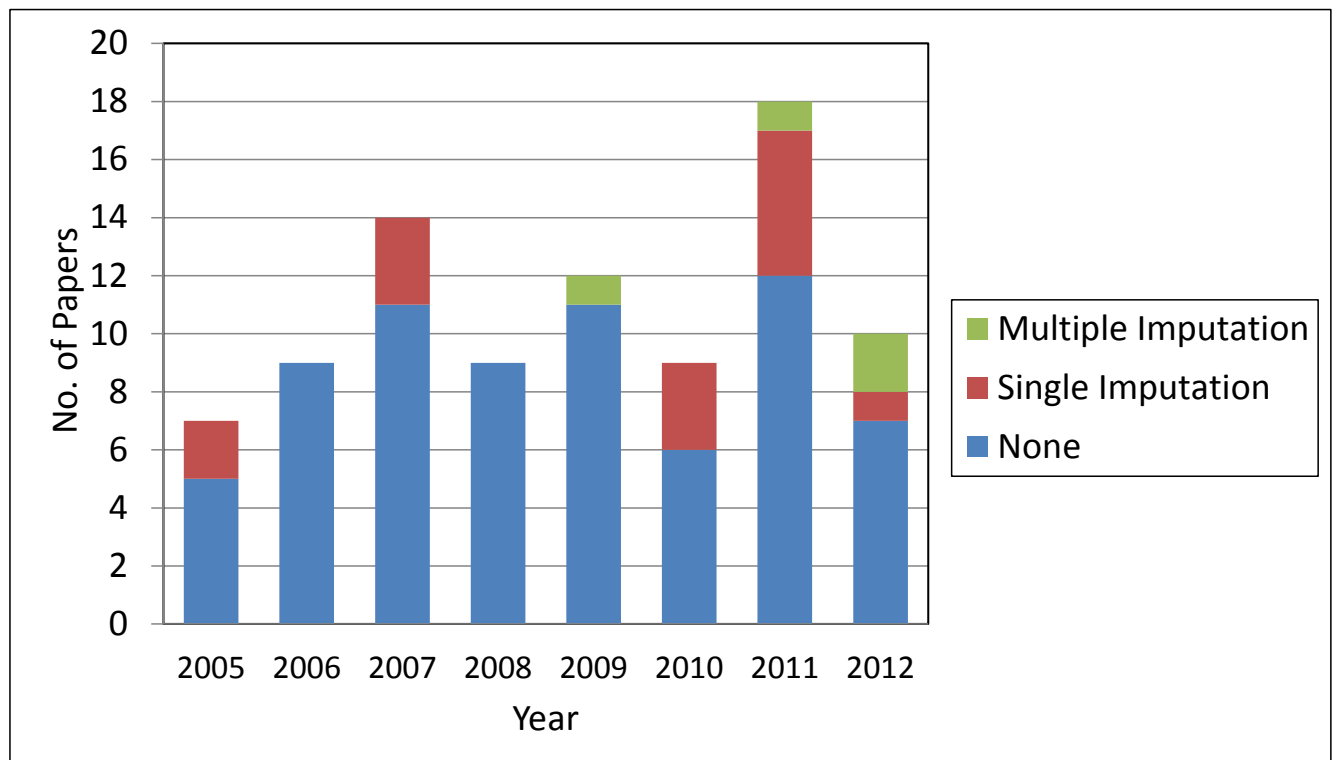
Scenario in Industry - 2012



Data source: BioMed Central

Part I: Missing Data

Year wise scenario of using imputation methods



Data source: BioMed Central

Presentation Plan

Part I: Missing data

- Why missing data matters?
- Key points in regulatory guidance
- Scenario in industry

Part II: Multiple Imputation Technique

- Steps
- SAS PROCs
- Methods available in PROC MI

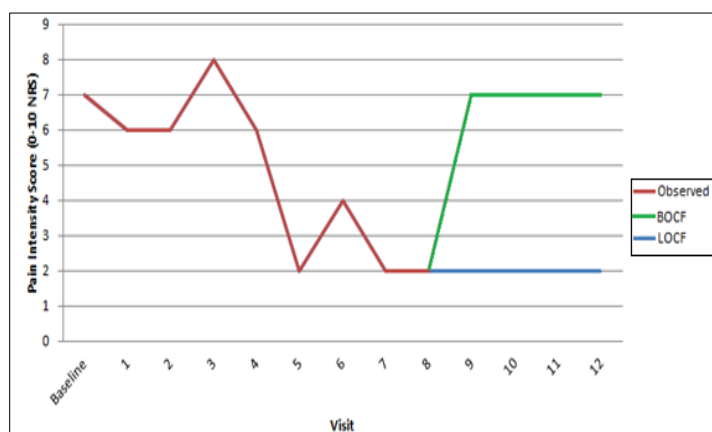
Part III: Comparison of methods

- Based on a simulation study

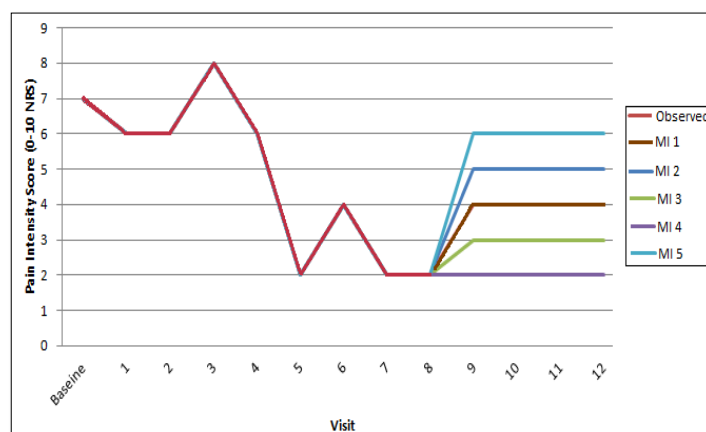
Part II: Multiple Imputation Technique

LOCF, BOCF and MI

Missing Pain Score Imputation –
LOCF and BOCF

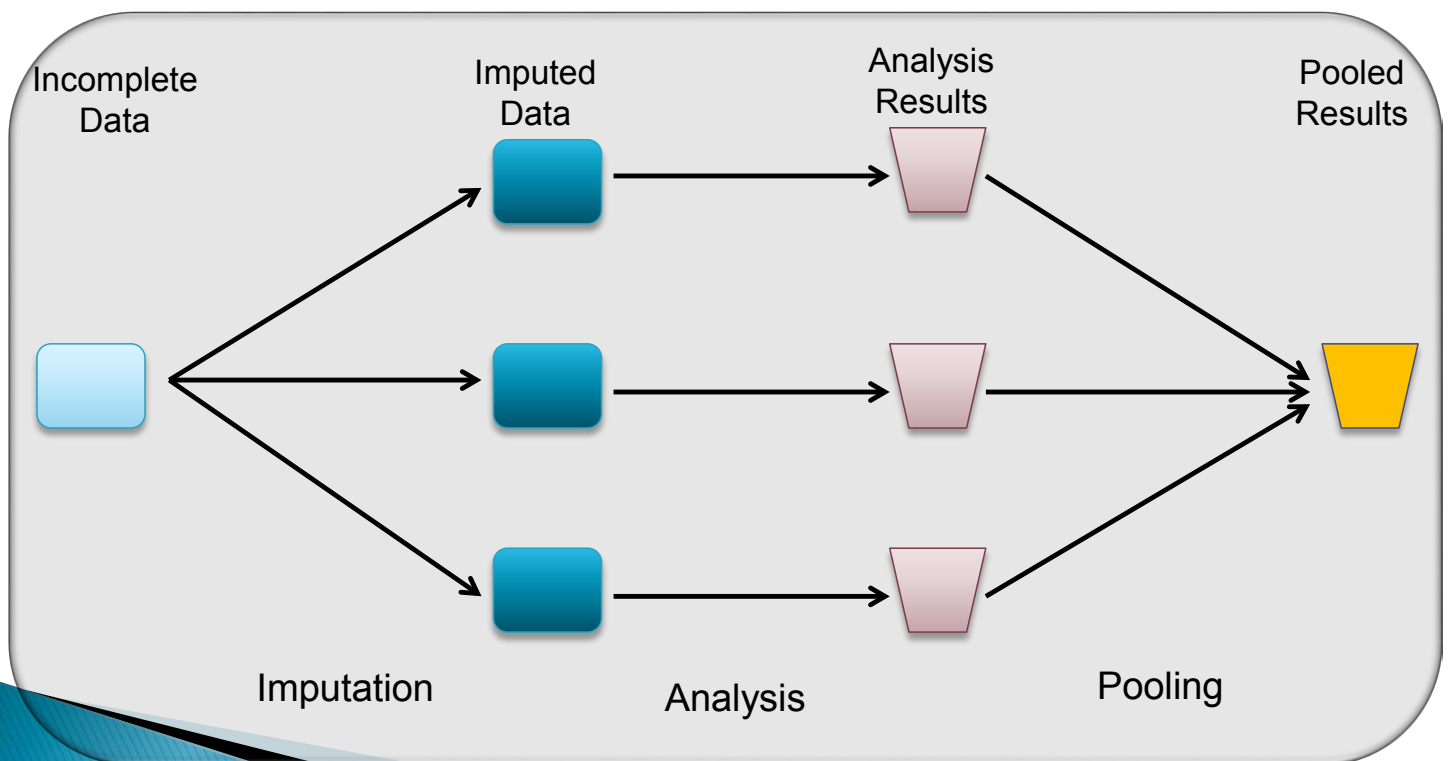


Missing Pain Score Imputation –
Multiple Imputation (MI)



Part II: Multiple Imputation Technique

Steps:



Part II: Multiple Imputation Technique

How the Missing value is filled?

Variable with missing values – Y

A related variable with no missing value – X

Missing Y values – Y_M

Non-missing Y values – Y_O

Use Y_O and corresponding X

Study relationship between X and Y

Draw Y_M randomly from $Y_M | X$ to complete the dataset

Repeat this for m times

Part II: Multiple Imputation Technique

SAS PROCS

- ❑ Imputation - **PROC MI**
- ❑ Analysis – Standard statistical methods (**PROC REG**, **PROC GLM**,
PROC MIXED, **PROC GENMOD**, etc.)
- ❑ Pooling - **PROC MIANALYZE**

For more details please visit

<http://support.sas.com/rnd/app/stat/papers/multipleimputation.pdf>

MI can be performed in other software also – R, Stata etc.

Part II: Multiple Imputation Technique

Methods Available in **PROC MI**

Missing Data Pattern	Type of Variables to be Imputed	Imputation Methods Recommended	SAS® Options in PROC MI
Monotone	Continuous	Regression method	MONOTONE REG
		Predicted mean matching	MONOTONE REGPMM
		Propensity score	MONOTONE PROPENSITY
Monotone	Categorical (Ordinal)	Logistic regression	MONOTONE LOGISTIC
	Categorical (Nominal)	Discriminant function method	MONOTONE DISCRIM
Arbitrary	Continuous	MCMC full data imputation	MCMC IMPUTE = FULL
		MCMC monotone data imputation	MCMC IMPUTE = MONOTONE

Part II: Multiple Imputation Technique

Missing Data Pattern

Monotone Pattern

v1	v2	v3	v4
X	X	X	.
X	X	X	.
X	X	.	.
X	.	.	.

Arbitrary Pattern

v1	v2	v3	v4
X	.	X	X
X	X	X	.
X	X	.	X
X	X	X	.

Part II: Multiple Imputation Technique

SAS example – missing pattern and percentage

```
proc mi data = pain nimpute = 0 ;  
  var ady trt01Pn pain;  
run;
```

Missing Data Patterns								
Group	ady	TRT01PN	pain	Freq	Percent	-----Group Means-----		
						ady	TRT01PN	pain
1	X	X	X	17935	87.54	37.573014	2.448341	3.085587
2	X	X	.	2552	12.46	50.322100	2.525862	.

Part II: Multiple Imputation Technique

	Rate of missing				
m	10%	20%	30%	50%	70%
3	0.9677	0.9375	0.9091	0.8571	0.8108
5	0.9804	0.9615	0.9434	0.9091	0.8772
10	0.9901	0.9804	0.9709	0.9524	0.9346
20	0.9950	0.9901	0.9852	0.9756	0.9662

SAS example – Imputation

```
proc mi data = pain nimpute = 10
  seed=314719001 out = miout;
  class ady trt01pn pain;
  monotone logistic;
  var ady trt01pn pain;
run;
```

Presentation Plan

Part I: Missing data

- Why missing data matters?
- Key points in regulatory guidance
- Scenario in industry

Part II: Multiple Imputation Technique

- Steps
- SAS PROCs
- Methods available in PROC MI

Part III: Comparison of methods

- Based on a simulation study

Part III: Comparison of methods

- ❑ Data were simulated – daily pain intensity score (0-10 NRS)
- ❑ 12 week study of test drug vs. placebo
- ❑ Primary endpoint – change from baseline in pain intensity score

→ Complete Data

- ❑ Data were deleted to establish missing completely at random

→ Incomplete Data

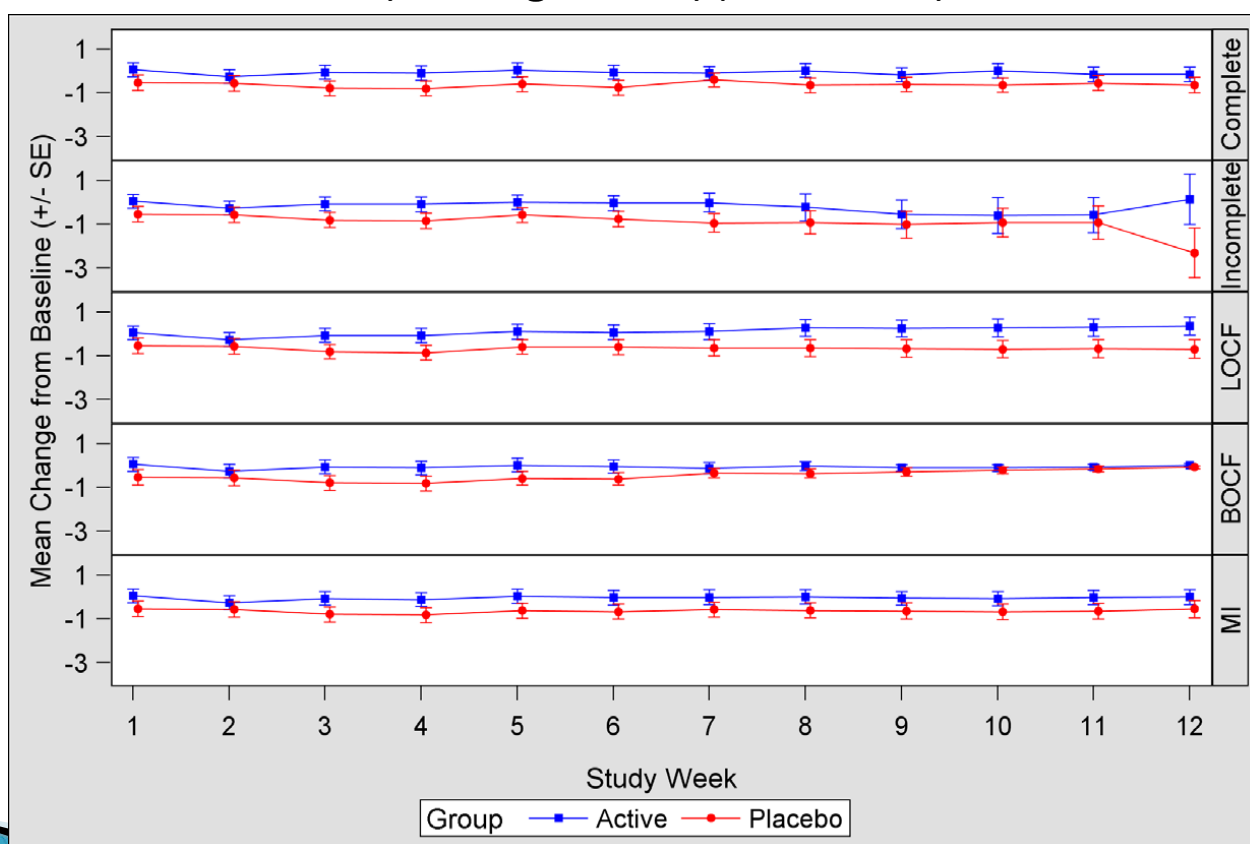
Part III: Comparison of methods

Change from baseline (CFB) in pain intensity score

Analysis Performed	Active		Placebo	
	Mean	SE	Mean	SE
Complete Data	-1.12 to 0.83	0.27 to 0.39	-1.17 to 1.05	0.29 to 0.39
Incomplete Data	-2.72 to 1.93	0.29 to 1.61	-3.83 to 2.71	0.29 to 1.52
LOCF	-1.12 to 1.00	0.30 to 0.51	-1.10 to 1.00	0.29 to 0.51
BOCF	-1.04 to 0.79	0.03 to 0.39	-1.10 to 0.98	0.04 to 0.40
MI	-1.11 to 0.87	0.29 to 0.41	-1.10 to 1.07	0.29 to 0.40

Part III: Comparison of methods

Mean CFB in weekly average of daily pain intensity score



Part III: Comparison of methods

Comparison of Imputation methods - significant LS Mean difference

Treatment Group	Comparison	Number of Cases p-value < alpha (0.1)
Active	Complete vs. Incomplete	27
	Complete vs. LOCF	37
	Complete vs. BOCF	33
	Complete vs. MI	0
Placebo	Complete vs. Incomplete	28
	Complete vs. LOCF	38
	Complete vs. BOCF	35
	Complete vs. MI	0

Summary

- ❑ Not to say that MI is always the best over LOCF and BOCF, but **rather than** MI should also be considered
- ❑ More research is warranted to further explore the use of multiple imputation in the setting of pain studies
- ❑ More work is required to analyze the data (not a case when m is modest)
- ❑ Detailed concept of missing data is required
- ❑ Should be very careful about the method to be used in PROC MI

References

- ▶ Rubin, D.B. (1987). Multiple imputation for nonresponse in surveys. New York: John Wiley & Sons, Inc.
- ▶ SAS/STAT® 13.1 User's Guide The MI Procedure, Chapter 61
- ▶ Amalia Karahalios, Laura Baglietto, John B Carlin, Dallas R English and Julie A Simpson¹. A review of the reporting and handling of missing data in cohort studies with repeated assessment of exposure measures. BMC Medical Research Methodology 2012

Questions?



Thank You!

Ashik.Chowdhury@Cytel.com

