TATA CONSULTANCY SERVICES
Experience certainty.

PhUSE SDE, Mumbai (03-Dec-2016)

Sensitivity analysis for missing data - A shift from the MAR assumption

Swati Rizhwani

Disclaimer

The views and opinions expressed in this presentation are solely of the authors and do not necessarily represent official policy or position of **T**ata **C**onsultancy **S**ervices

Discussion Points

- Introduction
- Missing Data Mechanisms
- Overview of analytic approaches to missing data
- Pattern Mixture Models to assess shift from MAR assumption
- Case Study
- References

Background:

The regulatory bodies recommend to address potential impact of anticipated missing data. The Statistical analysis plan of the study should discuss the appropriate methods to assess missing data patterns and to perform sensitivity analyses.

Prevention and Treatment of missing data:

Careful design and conduct to limit the amount of missing data

Careful attention to the assumptions about the nature of the missing data

Analysis that makes full use of information on all randomized participants

Missing Data Mechanisms

Let R be a “response” indicator having a value of 1 if Y is missing and 0 if Y is observed

- **M**issing **C**ompletely **a**t **R**andom (MCAR) : The data are *missing completely at random* (MCAR) if the probability that *Y* is missing does not depend on *X (other observed variables; covariates)* or on *Y* itself (Rubin 1976).

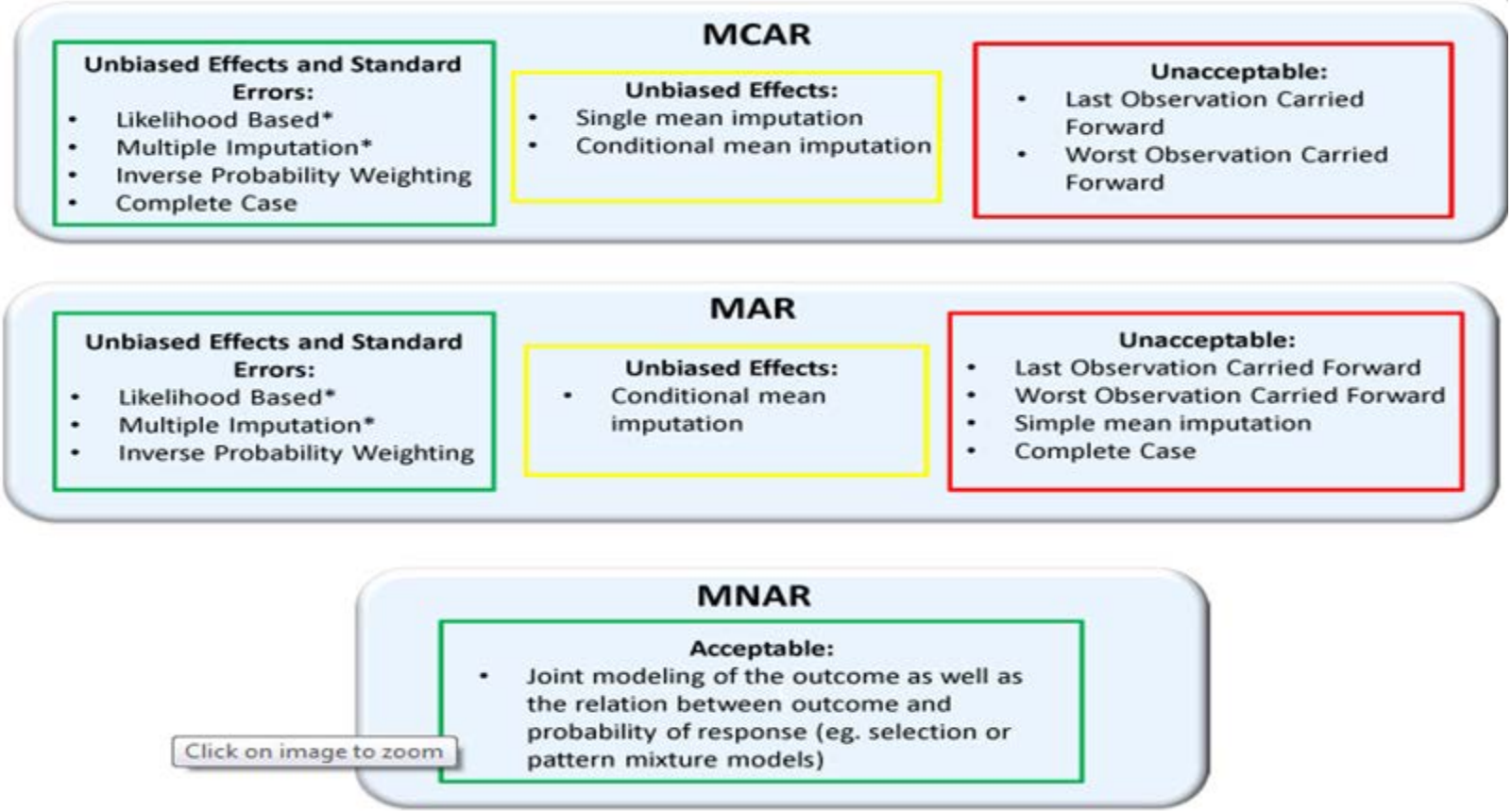
$$Pr(R = 1 | X, Y) = Pr(R = 1)$$

- **M**issing **a**t **R**andom (MAR) : The data are *missing at random* (MAR) if the probability that *Y* is missing does not depend on *Y*, once we control for *X (other observed variables; covariates)*.

$$Pr(R = 1 | X, Y) = Pr(R = 1 | X)$$

- **M**issing **N**ot **a**t **R**andom (MNAR) : The data are not *missing at random* (MNAR) if probability that Y is missing depends on Y itself i.e. $Pr(R = 1 | X, Y)$ cannot be reduced.

Overview of analytic approaches to missing data



Pattern Mixture Models to assess shift from MAR assumption

Pattern-mixture models (PMM) stratify incomplete data by the pattern of missing values and formulate distinct models within each stratum.

General PMM framework:

Notations:

Entire data matrix **Y : (**Y_{obs}**,** Y_{miss}**)** ; **X :** (observed) covariates ; **R :** matrix of indicators of missingness

PMMs decompose the joint probability of data and missingness as follows:

$$p(Y_{obs}, Y_{mis}, R|X) = p(R|X)\, p(Y_{obs}, Y_{mis}|\, R, X)$$
$$= p(R|X)\, p(Y_{obs}\,|\, R, X)\, p(Y_{mis}\,|\, Y_{obs}, R, X)$$

model for missing data conditioned on observed data within each pattern

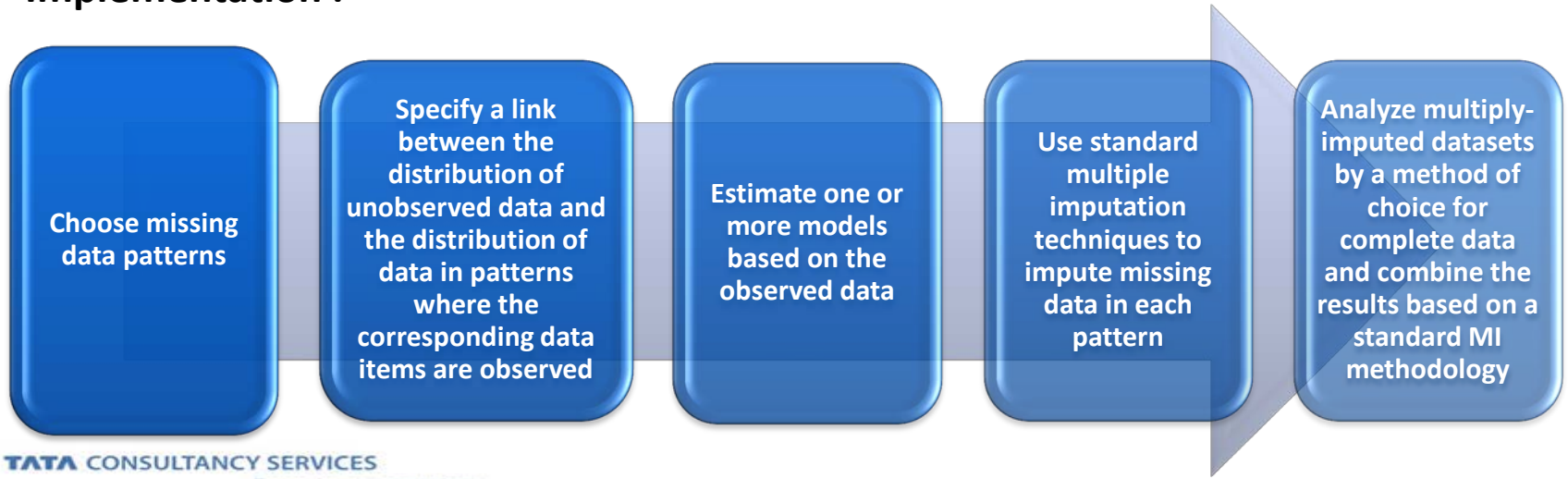
probability distribution of various missingness patterns

model for available data within each pattern

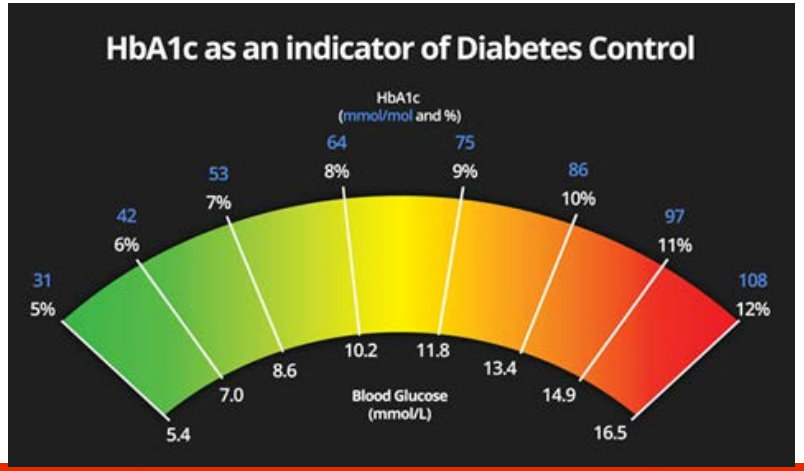
Pattern Mixture Models to assess shift from MAR assumption

- Pattern-mixture models are, by definition, under-identified because patterns with missing data typically have some parameters of the $p(Y_{obs}, Y_{mis} \mid R, X)$ model that cannot be estimated from data due to incomplete data within that pattern.
- So in order to estimate PMMs, one simply needs to explicitly impose some assumptions regarding the inestimable parameters, referred to in the PMM literature as “identifying restrictions”.

Implementation :



Background: HbA1c (glycated haemoglobin)



HbA1c	mmol/mol	%
Normal	**Below 42 mmol/mol**	**Below 6.0%**
Prediabetes	**42 to 47 mmol/mol**	**6.0% to 6.4%**
Diabetes	**48 mmol/mol or over**	**6.5% or over**

Objective:

- To show the non-inferiority of treatment A versus treatment B in terms of HbA1c change from baseline to Week 26 in patients with type 1 diabetes mellitus.

Primary efficacy endpoint:

- Change in HbA1c from baseline to Week 26 (Negative values preferred)

Analysis Plan:

- The primary efficacy variable will be analyzed using a mixed-effect model for repeated measures (MMRM) approach, under the missing at random (MAR) framework.
- Robustness of the primary efficacy analysis results in view of departure from the MAR assumption will be using tipping-point analysis based on the pattern mixture model approach.

Decision rule:

- The primary efficacy variable will be assessed for non-inferiority. Non-inferiority will be demonstrated if the upper bound of the two-sided 95% CI of the difference between treatment A and Treatment B is <0.3%.

Primary efficacy analysis :

Table 1:

Mixed-effect **M**odel for **R**epeated **M**easures (MMRM) approach (MAR framework)

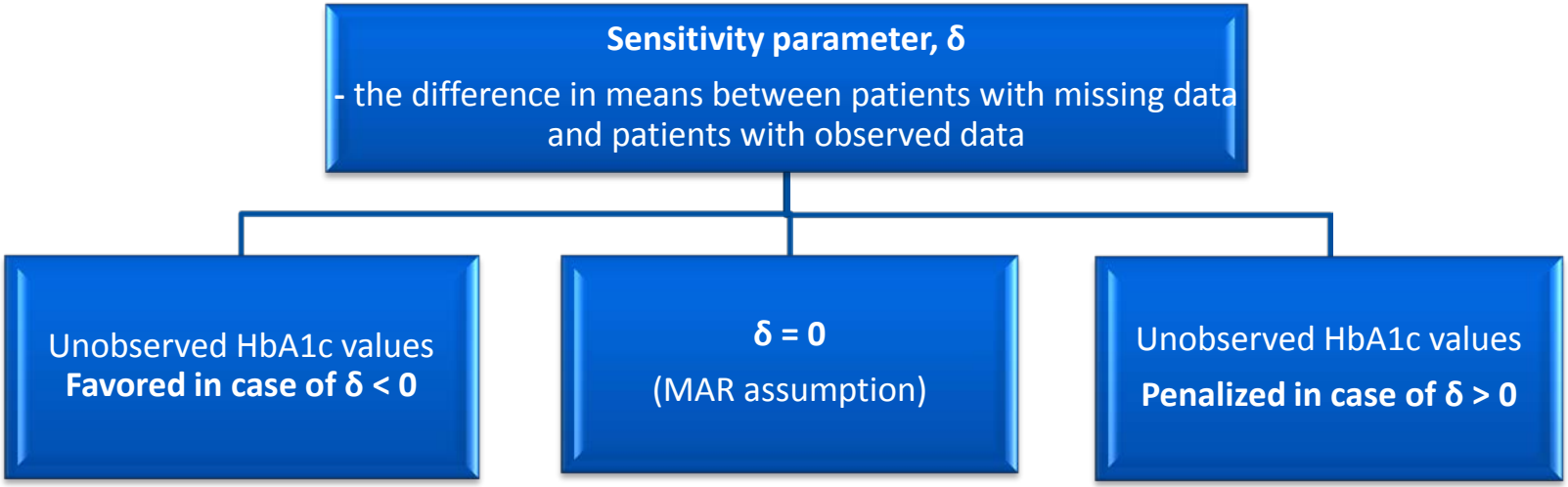
	Trt A	Trt B
LS Mean (SE)	-0.40 (0.05)	-0.44 (0.05)
95% CI	(-0.510 to -0.307)	(-0.533 to -0.374)
LS Mean difference (SE) Trt A vs Trt B	0.04 (0.073)	
95% CI	(-0.092 to 0.191)	

Step 1: Choose missing data patterns

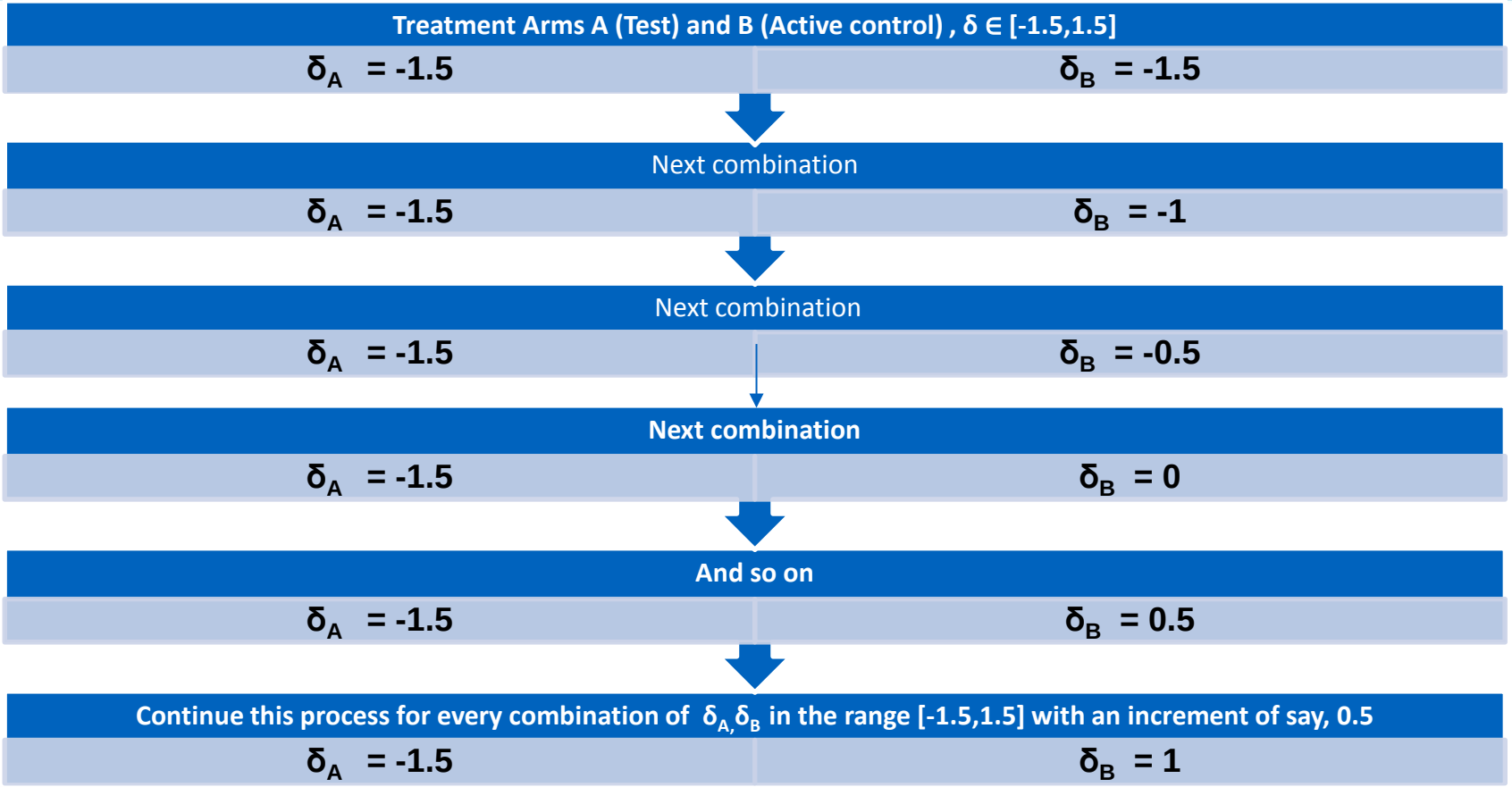
In PMMs, pattern represents a group of subjects that share some general common characteristics related to drop-out

HbA1c Missing Data Pattern	Trt A (N=223)	Trt B (N=224)
Patients with any missing data	18 (8.07%)	12 (5.36%)
Patients with missing value at baseline	0	0
Patients with available value at baseline but missing value at Week 12 and Week 26	8	6
Patients with available value at baseline and Week 12 but missing value at Week 26	5	3
Patients with available value at baseline and Week 26	210	215

Step 2: Specify a link between the distribution of unobserved data and the distribution of data in patterns where the corresponding data items are observed



Case Study


Treatment Arms A (Test) and B (Active control) , δ ∈ [-1.5,1.5]
δA = -1.5
δB = -1.5
Next combination
δA = -1.5
δB = -1
Next combination
δA = -1.5
δB = -0.5
Next combination
δA = -1.5
δB = 0
And so on
δA = -1.5
δB = 0.5
Continue this process for every combination of δA,δB in the range [-1.5,1.5] with an increment of say, 0.5
δA = -1.5
δB = 1

Step 3: Estimate one or more models based on the observed data

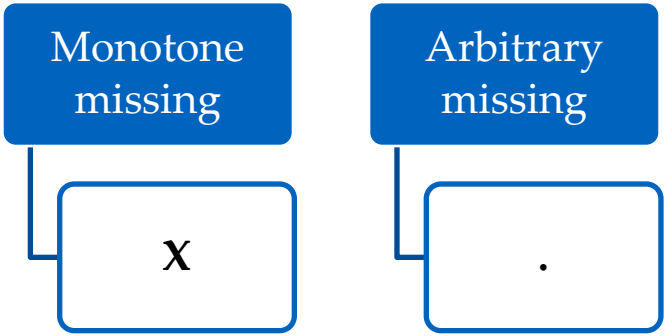
In our case, we consider two models on observed data,

1. to handle arbitrary missing values
2. to handle monotone missing values

USUBJID	ARM	AVISIT	AVAL
1	A	Baseline	6
1	A	Week 120	.
1	A	Week 260	6.5
2	B	Baseline	7
2	B	Week 120	7.2
2	B	Week 260	**X**
3	A	Baseline	8
3	A	Week 120	**X**
3	A	Week 260	**X**

Patterns identified for Multiple Imputation :

In PROC MI, pattern refers to a configuration or position of missing values when analysis variables are arranged in a certain order.



Step 4: According to the link function in (2), use standard multiple imputation techniques to impute missing data in each pattern with missing data based on draws from model(s) estimated in (3)

I. Multiple Imputation using Monte-Carlo Markov Chain (MCMC) :

```
proc mi data=efficacy out=_eff_nomiss nimpute=100 seed=12000 minimum=0 minmaxiter=10000 noprint;
    mcmc impute=monotone;
    var armn baseline week10 week20;
run;
```

	Imputation	SUBJID	ARMN	stratum_a	stratum_b	Baseline	Week10	Week20	miss_20
1	1		.	.	.	.	.	.	Y
2	1	1001001	1	2	1	7.9	7.2	7.8	
3	1	1001003	1	1	1	6.8	6.6	6.6	
4	1	1002001	1	2	1	7.8	7.5937756694	7.9	
21	1	2003005	1	1	1	6.9	6.9	6.4	
22	1	2004001	1	1	1	7.4	.	.	Y
23	1	2004002	1	1	1	7.4	7.5	7.1	
24	1	2004003	2	1	2	7.6	7.5	6.9	

II. Multiple Imputation using Monte-Carlo Markov Chain (MCMC) :

```
proc mi data=_eff_nomiss out=findat nimpute=1 seed=24000  noprint;
  by _imputation_;
  class armn  stratum_a  stratum_b ;
  monotone method=reg;
  var armn  stratum_a  stratum_b  baseline week10 week20;
run;
```

	Imputation	SUBJID	ARMN	stratum_a	stratum_b	Baseline	Week10	Week20	miss_20
1	1		.	.	.	.	.	.	Y
2	1	1001001	1	2	1	7.9	7.2	7.8	
3	1	1001003	1	1	1	6.8	6.6	6.6	
4	1	1002001	1	2	1	7.8	7.5937756694	7.9	
21	1	2003005	1	1	1	6.9	6.9	6.4	
22	1	2004001	1	1	1	7.4	6.298620754	6.7357531556	Y
23	1	2004002	1	1	1	7.4	7.5	7.1	
24	1	2004003	2	1	2	7.6	7.5	6.9	

Step 5: Analyze multiply-imputed datasets by a method of choice for complete data and combine the results based on a standard MI methodology

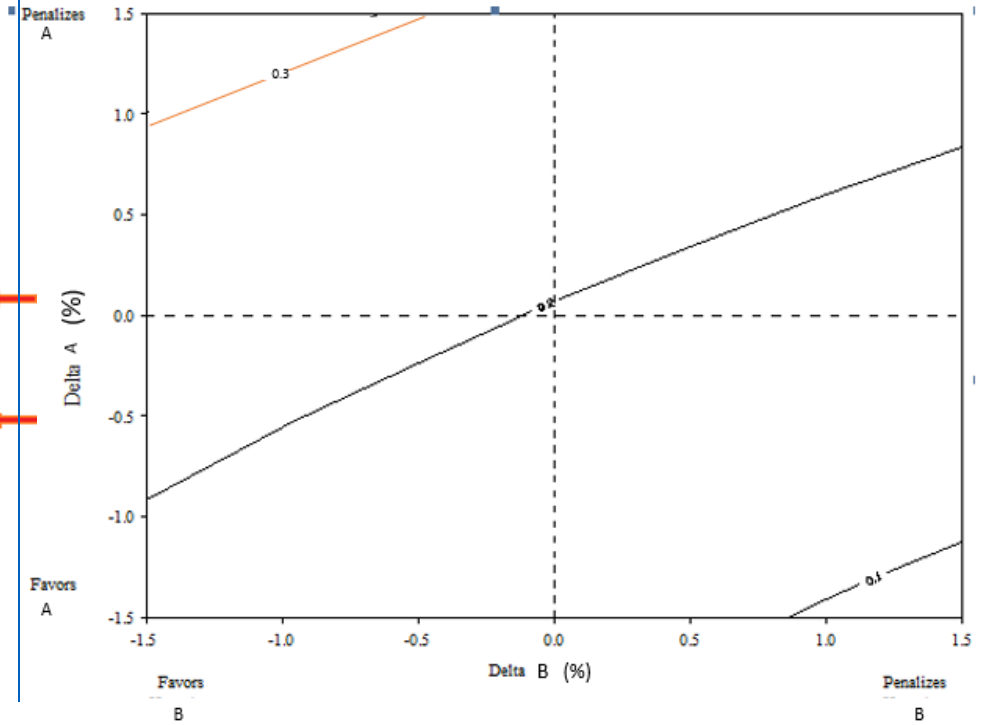
Post imputation, the complete datasets will be analyzed separately using an ANCOVA model. The results obtained will be combined using PROC MIANALYZE

```
proc mianalyze parms=_eff_ANCOVA_LSmean;
 by deltaA deltaB avisitn armn;
 modeleffects LSmean;
 ods output ParameterEstimates=_eff_Est_ANCOVA_LSmean;
 ods output VarianceInfo=_eff_Var_ANCOVA_LSmean;
run;
```

Dataset representing the set of combination of values for the shift parameters

LSMeandiff_SE	CI95	Delta_A	Delta_B
0.1196 (0.0740)	(-0.0254 to 0.2646)	0.3	-1.5
0.1298 (0.0743)	(-0.0159 to 0.2754)	0.5	-1.5
0.1399 (0.0748)	(-0.0066 to 0.2864)	0.7	-1.5
0.1501 (0.0753)	(0.0025 to 0.2976)	0.9	-1.5
0.1623 (0.0764)	(0.0126 to 0.3120)	1.5	-0.9
0.1562 (0.0761)	(0.0070 to 0.3055)	1.5	-0.7
0.1502 (0.0760)	(0.0013 to 0.2991)	1.5	-0.5
0.1441 (0.0759)	(-0.0046 to 0.2928)	1.5	-0.3
0.1380 (0.0758)	(-0.0106 to 0.2866)	1.5	-0.1

Graphical representation of the results : Contour plot



Conclusion:

- ✓ From Primary Efficacy analysis (Table 1) , we see that the upper bound of the two-sided 95% CI of the difference between treatment A (Test) and Treatment B (Active control), obtained by MMRM is 0.191 (<0.3%). Hence the hypothesis of non-inferiority of treatment A to treatment B cannot be rejected.

- ✓ Based on the sensitivity analysis using pattern mixture model approach to assess the impact of shift from MAR assumption, it is seen that 0.9 (delta_A, i.e. test treatment **penalized** to the extent **0.9** units) and -0.5 (delta_B, i.e. control treatment **favored** to the extent **0.5** units) are the points for treatment A and treatment B respectively, until where the null hypothesis of non-inferiority cannot be rejected i.e. the upper limit of the confidence interval for both of these points is within the range (<0.3) . This point is termed as tipping point; as beyond this point, non-inferiority doesn't hold.

The tipping point obtained is

Delta_A= 0.9

Delta_B= -0.5

References

- http://pharmasug.org/proceedings/2011/SP/PharmaSUG-2011-SP04.pdf
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3767219/
- http://www.theanalysisfactor.com/missing-data-mechanism/

Questions

TATA CONSULTANCY SERVICES

TATA

Thank You