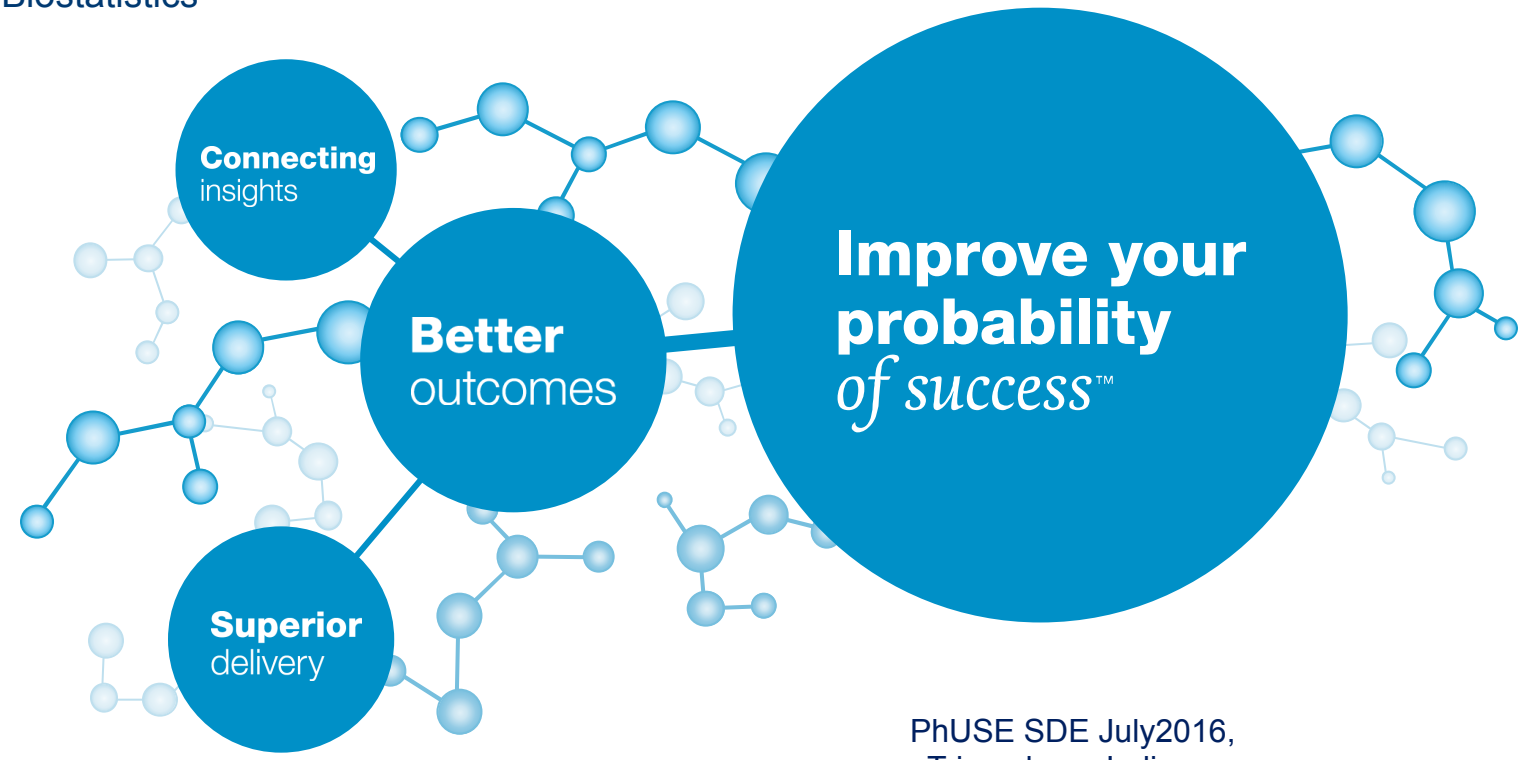
Statistical Planning for Missing Data in Clinical trials

Suresh Chenji
Manager Biostatistics



PhUSE SDE July2016,
Trivandrum, India

Abstract

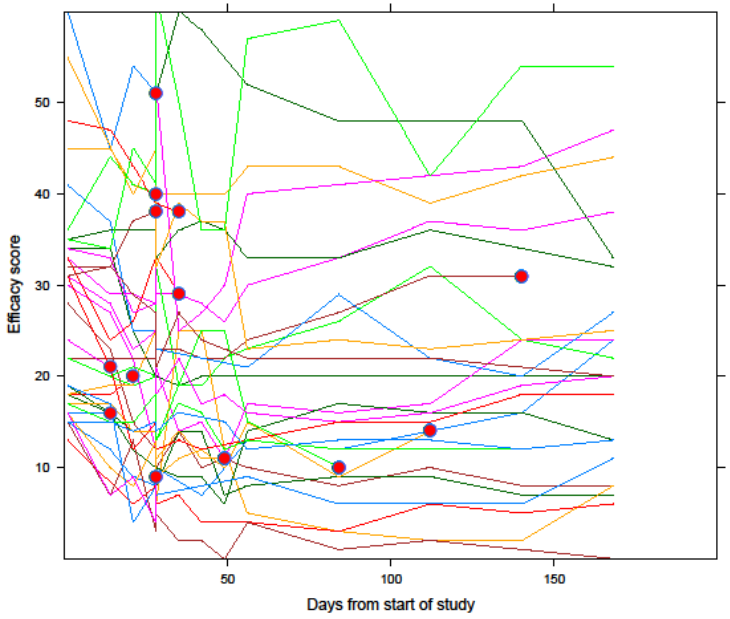
The problem of missing data has recently gained increased attention since the report "The prevention and treatment of missing data in clinical trials" published by the US National Research Council in 2010. Regulatory agencies have been concerned with single imputation approaches and requiring a more thorough investigation of missing data patterns and justifications for the method chosen to handle missing data, also requiring sensitivity analyses showing how different assumptions about missing data influence results.

Depending on clinical input regarding likely outcomes for withdrawals, for assumptions like missingness is independent of unobserved outcomes and missingness is dependent on the unobserved outcomes, various approaches - Direct likelihood approaches (Mixed effect Model Repeated Measures), Multiple imputation (pattern-mixture models framework) procedures, Control-based imputation for the experimental arm, Tipping point analysis using delta-adjustment for subjects on experimental arm etc., are to be considered for primary and sensitivity analyses.

Scientific approaches to handle missing data are becoming necessary to incorporate through trial protocols and statistical analysis plans, in order to draw valid conclusions regarding treatment efficacy from clinical trials.

Missing Data in Clinical Trials

- Early discontinuations (drop-outs) are the main source of missing data.
- What happens to subjects who leave the clinical trial prematurely?
- How does the missing info affect our judgment about the experimental treatment ?



- Reduced power
- Potential for bias and increase in Type I error
  - Introduce bias in estimated treatment contrast and the confidence interval around it

Missing Data Mechanisms

- **Missing Completely at Random (MCAR)**: the probability of an observation being missing does not depend on observed or unobserved measurements - Strongest assumption. If this is the case, analysis of the completers will be unbiased and good enough.
  - Example: Lab sample accidentally broken, Patient moves to another city for non-health reasons
- **Missing at Random (MAR)**: The reason for data being missing may depend on observed data (trajectory), but not on the unobserved missing data.
  - Example: Subjects withdrawn due to observed high HbA1c values (lack of efficacy)
- **Missing Not At Random (MNAR):** When observations are neither MCAR nor MAR, they are classified as Missing Not At Random (MNAR), i.e. the probability of an observation being missing depends on unobserved measurements.
  - Example: study evaluating treatments to reduce cocaine use in which the outcome is drug level from a urine drug test measured every Monday morning. Participants who use cocaine over the weekend and do not show up for their urine test would be expected to have higher levels of cocaine metabolites. Thus, the likelihood of the data being missing is directly related to the unobserved cocaine level.

(Non-ignorable)

Common Ways to Deal with Missing Data

Imputation + ANCOVA

$$y = \begin{pmatrix} y_{1K} \\ y_{2K} \\ \vdots \\ y_{NK} \end{pmatrix}$$

- Missing values are imputed first
  - Single imputation, e.g., LOCF, BOCF
  - Multiple Imputation (MI)
- Imputed data analyzed at the primary timepoint with ANCOVA
  - ANCOVA does not take into account data at other timepoints and correlation between timepoints of the same patient.

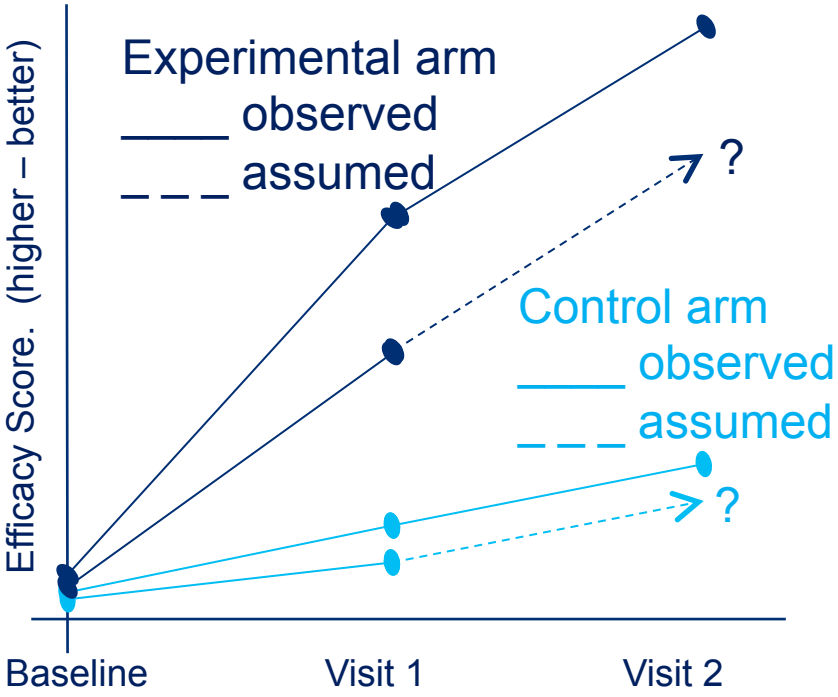
MMRM (Mixed Model with Repeated Measures)

$$y = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1K} \\ y_{21} & y_{22} & \cdots & y_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NK} \end{pmatrix}$$

- Missing values are not imputed but all subjects and all available timepoints are included in analysis.
- Correlation between data at all timepoints is modeled.
  - Partial data (prior to drop-out) is used when estimating the variance-covariance matrix.
- Inference at a particular timepoint is obtained with a contrast from the overall model, which includes all visits.
  - MMRM-based inference takes the correlation into account.
  - Can be implemented in SAS using PROC MIXED

MMRM Assumptions

- MMRM inference is based on Missing at Random (MAR) assumptions.
- MAR: missingness is independent of unobserved data after accounting for observed data in the model.



MMRM assumes:

- Drop-outs are similar to completers conditional on observed data included in the model
- Thus drop-outs may continue to reap benefits from treatment (just like completers) as if they were still on treatment

This may not be clinically reasonable in some cases

MAR Applicability

- Treatment of symptoms vs. treatment of underlying disease
- In symptomatic treatment trials, patients that cannot continue with treatment (e.g., due to intolerance), should not be assumed to derive any future benefit from it after discontinuation.
- However, not all drop-outs are the same – consider reason for discontinuation
  - Patients discontinuing due to adverse events (AEs) or lack of efficacy – clearly no future benefit from symptomatic treatment, if disease not affected
  - Patients discontinuing due to administrative reasons, lost to follow-up - potentially could continue with treatment, so possibly some future benefit if stayed on treatment

Multiple Imputation (MI)

- Standard MI relies on the assumption of data missing at random (MAR)
  - Missingness can be 'explained' by observed data
  - Missingness does not depend on unobserved data after accounting for observed factors
- Often withdrawal is indeed highly associated with observed data, e.g., treatment, baseline, assessment just before withdrawal, slope of change, etc.
- Imputed values are generated by an imputation model estimated based on observed data
- Each missing value is replaced with multiple imputed values
  - Multiple imputed datasets are produced
  - Each dataset has imputed values generated with an imputation model having slightly different parameter settings
- Each imputed dataset is analyzed separately using an analysis model, and then the results are combined using Rubin's rule taking into account uncertainty (variability) of different imputations
  - Analysis model can be different from imputation model
- MI can be used for MNAR imputations

Missing Patterns: Monotone vs. Non-monotone

- Pattern affects the distribution of the missing data matrix and affects how to handle imputations

- Monotone: Missing data always occur at the end of data records. i.e., for each record, values of $Y_1,\ldots, Y_{j-1}$ are non-missing; once value of Y_j is missing, all remaining variables $Y_{j+1},\ldots, Y_p$ are missing
- This allows to use univariate imputation models – impute one variable at a time, sequentially from left to right

Monotone: permanent discontinuation

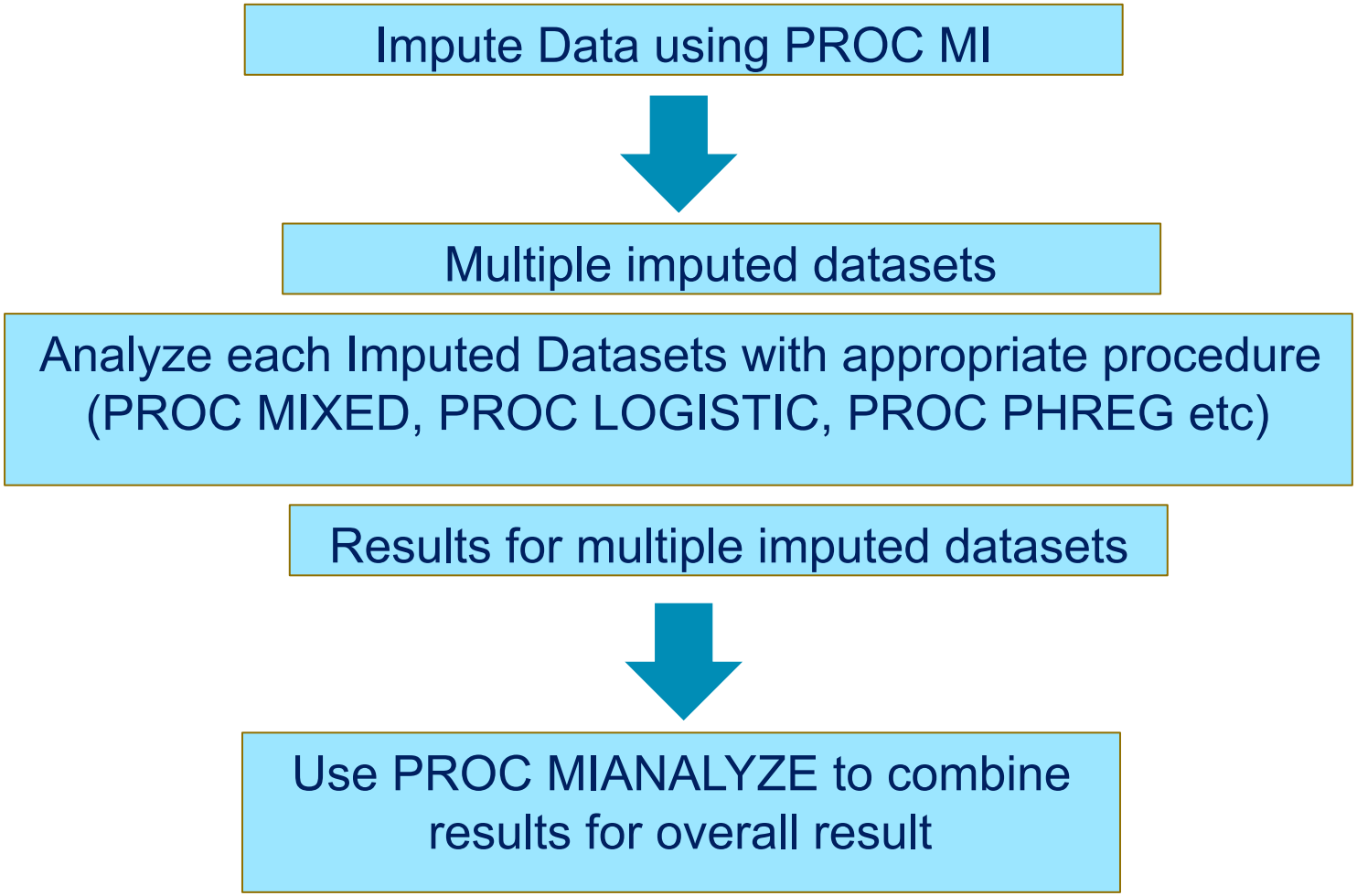
Y_1	Y_2	Y_3	Y_4
X	X	X	?
X	X	?	?
X	?	?	?

Non-monotone: MI requires all data to be imputed together treating these data as a multivariate response by creating a multivariate imputation model with a joint distribution of all missing and relevant explanatory variables

Non-monotone: Intermittent missing

Y_1	Y_2	Y_3	Y_4
X	X	?	X
X	X	?	?
X	?	X	?

Multiple Imputation Steps



Imputation Methods in PROC MI

Missing Data Pattern	Variable Type	Method
Monotone (MONOTONE statement)	Continuous	Linear regression Predictive mean matching Propensity score
	Binary/Ordinal	Logistic regression
	Nominal	Discriminant function Multinomial logistic regression
Non-monotone (MCMC statement)	Continuous	With continuous covariates: MCMC monotone method MCMC full-data imputation
(FCS statement)	Continuous	With mixed covariates: FCS regression FCS predictive mean matching
	Binary/ordinal	FCS logistic regression
	Nominal	FCS discriminant function FCS multinomial logistic regression

For Missing Not At Random (MNAR) assumption:
MNAR statement available with MODEL and ADJUST options –new in SAS v9.4
This statement is applicable only if it is used along with a MONOTONE statement or an FCS statement.

Placebo (Control) – Based Imputation

Missing Not At Random (MNAR)

- Subjects who discontinued from the experimental treatment will be assumed to follow a similar outcome trajectory as subjects from the placebo (control) arm

- Subjects who discontinued from placebo (control) treatment are modeled as completers within their own arm (MAR within control arm)
- One imputation model of placebo/control arm outcomes will be used to impute missing values for all discontinued subjects regardless of their randomized treatment

- Treatment effect will not be larger than that of placebo/control treatment and trial participation effect after subject discontinuation

Delta-adjusting Pattern Imputation

Missing Not At Random (MNAR)

- Subjects from the experimental treatment arm who discontinue at a given time-point would have, on average, their unobserved efficacy score worse by some amount **δ** compared to the observed efficacy score of subjects that continue to the next time point.

- Subjects who discontinue from the control arm would exhibit the same evolution of the disease as control subjects that stay in the study.

- Sensitivity analysis can be performed for a range of δ to find a **"tipping point"** – value of δ at which study conclusions change (from favorable to non-significant with respect to treatment efficacy).

Hypothetical Example Dataset Results

Change in HbA1c after 12 months

Assumption	Imputation Tool/Method	Estimate (CI) Active-Control	P-value
MAR	ANCOVA/LOCF	-1.39 (-1.64 ; -1.14)	<.0001
MAR	MMRM (PROC MIXED)	-1.49 (-1.74 ; -1.24)	0.0084
MAR	PROC MI/MONOTONE REG	-1.48 (-1.74 ; -1.22)	0.0079
MNAR: Control-Based	PROC MI (v9.4)/MONOTONE REG with MODEL	-1.41 (-1.68 ; -1.14)	0.0182
MNAR: Delta=0.2 for TRT, All Visits	PROC MI (v9.4)/MONOTONE REG with ADJUST	-1.45 (-1.71 ; -1.19)	0.0084
MNAR: Delta=0.4 for TRT, All Visits	PROC MI (v9.4)/MONOTONE REG with ADJUST	-1.14 (-1.39 ; -0.89)	0.0424
MNAR: Delta=0.6 for TRT, All Visits	PROC MI (v9.4)/MONOTONE REG with ADJUST	-1.01 (-1.27 ; -0.75)	0.0642

Summary

- A missing at random (MAR) approach is often justifiable as a primary analysis although likely biased to some degree.
- Multiple imputation (MI) can be used in the framework of pattern-mixture models (PMMs) to implement a variety of sensitivity analyses that are useful across many indications and treatments in stress-testing the primary analysis.
- Sensitivity analyses need to be pre-planned and pre-specified in protocol/SAP.
- Some level of exploratory sensitivity analyses may sometimes be warranted if pre-planned analyses do not address unexpected patterns of missing data.

PROC MI - Syntax

```
PROC MI < options > ;
BY variables ;
CLASS variables ;
EM < options > ;
FREQ variable ;
MCMC < options > ;
MONOTONE < options > ;
TRANSFORM transform ( variables < / options >) < . . .
   transform ( variables < / options >)
> ;
VAR variables ;
```

Expectation-Maximization Method
Single imputation with maximum likelihood estimates from a multivariate normal model (typically used as a starting point for MCMC).

Monte Carlo Markov Chain Method for arbitrary missing pattern

Methods/models for imputing data with monotone missing pattern

Transform variables before modeling (e.g., to deal with non-normality)

Variables to impute and/or include in the imputation model
Order is very important !!!

PROC MIANALYZE – Syntax and Input Datasets

```
PROC MIANALYZE < options > ;
   BY variables ;
   CLASS variables ;
   MODELEFFECTS effects ;
   < label: > TEST equat ion1 < , . . . , < equat ionk > > < / opt ions > ;
   STDERR variables ;
```

- **PROC MIANALYZE <Input Data Set Option> =<Input Data Set Name>**
- <Input Data Set Option>:
  - DATA= specifies COV, CORR, or EST type data set
  - DATA= specifies data set for parameter estimates and standard errors
  - PARMS= specifies data set for parameter estimates
  - PARMINFO= specifies data set for parameter information
  - COVB= specifies data set for covariance matrices
  - XPXI= specifies data set for $(X'X)^{-1}$ matrices

- **STDERR** statement lists the standard errors associated with the effects in the MODELEFFECTS statement when both parameter estimates and standard errors are saved as variables in the same DATA= data set. The STDERR statement can be used only when each effect in the MODELEFFECTS statement is a continuous variable by itself.

- SAS 9.2 User's Guide, PROC MIANALYZE, has examples for analysis with many different SAS procedures and how to feed their results into PROC MIANALYZE.

References

- European Medicines Agency. Guideline on Missing Data in Confirmatory Clinical Trials. July 2010.
- National Research Council (2010). The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academic Press.
- Carpenter J and Kenward M, 2008, Missing data in randomised controlled trials – a practical guide, Birmingham: National Health Service Co-ordinating Center for Research Methodology, www.missingdata.org.uk
- Molenberghs, G., Kenward, M. G., Missing Data in Clinical Studies. Wiley, 2007.
- Berglund and Heeringa. Multiple Imputation of Missing Data Using SAS.
- Ratitch B, O'Kelly M, Tosiello R (2013) Missing Data in Clinical Trials: from Clinical Assumptions to Statistical Analysis using Pattern Mixture Models. Pharmaceutical Statistics
- Siddiqui O, Hung HMJ and O'Neill, R, MMRM vs. LOCF: a comprehensive comparison based on simulation study and 25 datasets, Journal of Biopharmaceutical Statistics, 19, 227-246
- Ratitch, B. and O'Kelly, M. (2011), "Implementation of Pattern-Mixture Models Using Standard SAS/ STAT Procedures," in Proceedings of PharmaSUG 2011 (Pharmaceutical Industry SAS Users Group), SP04, Nashville

THANK YOU

QUINTILES®