



SP05:

Facing Your Fear of the Unknown – Handling Missing Data

PhUSE 2015 – Vienna
Amanda Reilly
ICON plc, Dublin, Ireland

- What?
 - What are missing data?
- Why?
 - Why not just ignore missing data?
- When?
 - When should missing data be acknowledged?
- How?
 - How should missing data be handled?

What?

WHAT ARE MISSING DATA?

What are missing data?

- Missing data are values that
 - are not available
 - would be meaningful
- Ideally, data would be complete (rarely the case)
- Common, serious & complex issue for clinical trials
 - Can limit conclusions
 - Can camouflage results

Why?

WHY NOT IGNORE MISSING DATA?

Why not ignore missing data?

- Subject considerably worsened
 - Skipped scheduled visit / withdrew from trial
 - Had they attended / continued in the trial; those measurements may have impacted their results
- Subject didn't answer a particular question
 - Embarrassed/afraid to admit something of interest?

When?

WHEN SHOULD MISSING DATA BE ACKNOWLEDGED?

When to acknowledge missing data?

- The power of a trial
 - Ability to detect difference between treatment & control groups (for an effective treatment)
 - Avoid incorrect conclusion of ineffective treatment (type II error)
 - Increases along with its sample size
- Bias
 - Inclination to be in favour of one treatment group
 - Ignoring uncertainty of missing values when S.Es are estimated

When to acknowledge missing data?

- Does this missing data matter?
- Considerations:
 - Power of the trial
 - Bias in treatment effect estimation
- European Medicines Agency (EMA); 1995
 - Committee for Medicinal Products for Human Use (CHMP)
 - ICH Topic E 9 - Statistical Principles for Clinical Trials (1998)
 - Points to Consider on Missing Data (2001)
 - Guideline on Missing Data in Confirmatory Clinical Trials (2010)

When to acknowledge missing data?

- MCAR
 - Missing Completely at Random
- MAR
 - Missing at Random
- MNAR
 - Missing Not at Random

When to acknowledge missing data?

- MCAR
 - Missing Completely at Random
- MAR
 - Missing at Random
- MNAR
 - Missing Not at Random

When to acknowledge missing data?

- MCAR
 - Missing Completely at Random
 - CHMP: “probability of an observation being missing does not depend on observed or unobserved measurements” for the dependent variable (conditional on covariates in model)
 - Presence /Absence unrelated to value that may have been observed

When to acknowledge missing data?

- MCAR
 - Missing Completely at Random
- MAR
 - Missing at Random
- MNAR
 - Missing Not at Random

When to acknowledge missing data?

- MAR
 - Missing at Random
 - CHMP: “conditional on the covariates in the model, the probability of an observation being missing depends only on observed measurements”
 - Implies that post-dropout behaviour can be predicted from pre-dropout observed behaviours
 - response can be estimated without bias

When to acknowledge missing data?

- MCAR
 - Missing Completely at Random
- MAR
 - Missing at Random
- MNAR
 - Missing Not at Random

When to acknowledge missing data?

- MNAR
 - Missing Not at Random
 - neither MCAR nor MAR
 - CHMP: “the probability of an observation being missing depends on unobserved measurements”, where “the value of the unobserved responses depends on information not available for the analysis”
 - unobserved observations not predicted without bias

How?

HOW SHOULD MISSING DATA BE HANDLED?

Handling missing data

- Only one correct way....

Handling missing data

- Only one correct way....

Don't have any!!!

- Only one correct way....

Don't have any!!!

- NB: Minimize missing data (protocol)
- CHMP Guidelines (2010): aim for complete capture
- ICH E9: missing data may be compensated for
- Note number, timing pattern, reason for missing values

Handling missing data

- Deletion methods
 - Most basic ways to handle missing data
- Imputation
 - Missing values are replaced with a value
 - Available data is used to make best estimates
- Modelling

Handling missing data

- Deletion methods
- Imputation
- Modelling

Deletion Methods

- Most basic and common techniques
- Observations with missing data are excluded
- Example: Listwise deletion
 - entire record is removed if it contains any missing value
- reduced sample size -> reduced statistical power
- biased estimates

Handling missing data

- Deletion methods
- Imputation
- Modelling

- Represent missing values with 'imputed' values
- Theory constantly developing
- Considerations
 - Type of data (numerical / binary / categorical)
 - Structure/Design of data
 - should respect the hierarchical structure of the data

Method 1: LOCF

- Widely used in the past
- Suitable for longitudinal studies
- Carry forward most recent observed post-baseline value (if one exists)
- Simple to understand and apply
- Biased estimates
- Now strongly discredited and discouraged

Method 2: Best/Worst Case

- Binary outcome data
- Worst case for test drug
- Best case for test drug
- Upper and lower limits of treatment effect

Method 3: Simple Imputation

- Missing value replaced with some single value
 - LOCF
 - Best/worst case (binary)
 - Mean of non-missing values (numerical)
 - Predicted values from regression analysis
- Issues:
 - Biased parameter estimates
 - Small standard errors => Type 1 error issues
 - Doesn't accurately reflect uncertainty of missing value

Method 4: Multiple Imputation

- Adds variability: reflects uncertainty of unknown values
- Each missing value has distribution of plausible values (based on observed data)
- Uncertainty of missing value reflected in the distribution
- Assumes data MAR
- PROC MI
 - Creates M plausible value to replace each missing value
 - Random sample reflects uncertainty of the missing value
 - M datasets analysed and results combined (PROC MIANALYZE)

Handling missing data

- Deletion methods
- Imputation
- Modelling

- Linear Mixed Models
 - MMARM; Mixed-Effect Model Repeated Measures
- Marginal / Random Effects Approaches
 - MAR: unbiased treatment effect estimate
- Pattern Mixture Model

- Only correct method: no missing data!
 - Use methods with different assumptions
 - Check for consistency across results
- Treatment vs. Study discontinuation

- Missing data: unavailable, would be meaningful
- Considerations:
 - power
 - Bias
 - CHMP Guideline on Missing Data (2010)
 - Type of missing data (MCAR / MAR / MNAR)
- Handling missing data
 - Deletion Methods
 - Single Imputation / Multiple Imputation
 - Modelling Approaches

Further Reading

Paper SP05 for syntax

Prevention: Little et al. “Prevention and Treatment of Missing Data in Clinical Trials”. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3771340/>

LOCF #1: Jain and Pandya “LOCF: It’s More Than Just Carrying Forward”.
<http://www.lexjansen.com/nesug/nesug09/ph/PH03.pdf>

LOCF #2: Ting, N. and Brailey, A. (2012). “Why do people use locf? or why not?”.
<http://www.doc88.com/p-517461081967.html>

CHMP “ICH Topic E 9 - Statistical principles for clinical trials” (1998), “Guideline on Missing Data in Confirmatory Clinical Trials” (2010). <http://www.ema.europa.eu>

Pattern Mixture Models: Ratitch, B. and OKelly, M. (2011). “Implementation of pattern mixture models using standard sas/stat procedures”, PharmaSUG2011.

Questions?