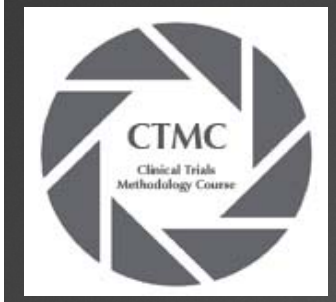

What is Missing from my Missing Data Plan?

RENEE' H. MARTIN, PHD

SHARON D. YEATTS, PHD

MEDICAL UNIVERSITY OF SOUTH CAROLINA

Supported by NINDS R25NS088248 U01NS077352

Disclosures

- Consultant fees from Genentech related to Role on PRISMS Trial Steering Committee
- Institutional fees from Bard related to DSMB service

Supported by NINDS R25NS088248 U01NS077352

Examples: Weight Loss

about the effects of low-carbohydrate diets. It is unfortunate, however, that so much effort must be devoted to evaluating the implications of missing observations when a seemingly simple effort to obtain study weights according to the follow-up protocol would probably have been successful with most participants. Complete evaluation of enrolled patients, irrespective of their adherence to study therapy, deserves wider recognition as an important part of good clinical-trials practice.

tion from
analyses fi
between tr
ization. Hc
analysis, a
tion must b
In these tw
not follow
continued the study therapy.

Unfortu
itive conclu
ty of carbo
in these tri
participant
In
dom
diets
et al.
mon
high
mon
132 s
by Foster a
cent) completed 12 months of follow-up, and 79 of
132 subjects (60 percent) completed 6 months of
follow-up in the study by Samaha et al. The defini-
[the weight in kilograms divided by the square of
the height in meters] of at least 35).

Ware JH. Interpreting incomplete data in studies of diet and weight loss. N Engl J Med 2003;348:2136-7.

Supported by NINDS R25NS088248 U01NS077352

Intention-to-Treat Principle

- Analysis should include all randomized subjects in the group to which they were randomly assigned, regardless of
  - Eligibility violations
  - Adherence to assigned treatment
  - Treatment received
  - Protocol violations

- Missing data complicates the statistical analysis under the ITT principle. Typical procedures in statistical programs will exclude from analysis subjects for whom one of the specified variables is missing.

Supported by NINDS R25NS088248 U01NS077352

Common Reasons for Missing Data

- Reasons
  - Dropouts or death
  - Missed visit, or visit conducted outside of specified window
  - Non-response, to a single item in assessment, or to an entire assessment
  - Site error
- Definition (Little et al, NEJM 2012, 367: 1355-1360)
  - ”values that are not available and that would be meaningful for analysis if they were observed”

Supported by NINDS R25NS088248 U01NS077352

Impact of Missing Data

- Loss of information → Loss of power
  - Available sample size lower than intended
- Bias
  - Missingness may be associated with response to treatment
  - Completers subgroup is not a random sample of the original

Supported by NINDS R25NS088248 U01NS077352

Constructing a Missing Data Plan

- Minimize missing data
  - By design
  - During implementation
- Analysis Strategies

Supported by NINDS R25NS088248 U01NS077352

Minimize Missing Data

Table 1. Eight Ideas for Limiting Missing Data in the Design of Clinical Trials.

Target a population that is not adequately served by current treatments and hence has an incentive to remain in the study.
Include a run-in period in which all patients are assigned to the active treatment, after which only those who tolerated and adhered to the therapy undergo randomization.
Allow a flexible treatment regimen that accommodates individual differences in efficacy and side effects in order to reduce the dropout rate because of a lack of efficacy or tolerability.
Consider add-on designs, in which a study treatment is added to an existing treatment, typically with a different mechanism of action known to be effective in previous studies.
Shorten the follow-up period for the primary outcome.
Allow the use of rescue medications that are designated as components of a treatment regimen in the study protocol.
For assessment of long-term efficacy (which is associated with an increased dropout rate), consider a randomized withdrawal design, in which only participants who have already received a study treatment without dropping out undergo randomization to continue to receive the treatment or switch to placebo.
Avoid outcome measures that are likely to lead to substantial missing data. In some cases, it may be appropriate to consider the time until the use of a rescue treatment as an outcome measure or the discontinuation of a study treatment as a form of treatment failure.

Little et al, NEJM 2012, 367: 1355-1360;

Supported by NINDS R25NS088248 U01NS077352

Minimize Missing Data

Table 2. Eight Ideas for Limiting Missing Data in the Conduct of Clinical Trials.

Select investigators who have a good track record with respect to enrolling and following participants and collecting complete data in previous trials.
Set acceptable target rates for missing data and monitor the progress of the trial with respect to these targets.
Provide monetary and nonmonetary incentives to investigators and participants for completeness of data collection, as long as they meet rigorous ethical requirements.[15,16]
Limit the burden and inconvenience of data collection on the participants, and make the study experience as positive as possible.
Provide continued access to effective treatments after the trial, before treatment approval.
Train investigators and study staff that keeping participants in the trial until the end is important, regardless of whether they continue to receive the assigned treatment. Convey this information to study participants.
Collect information from participants regarding the likelihood that they will drop out, and use this information to attempt to reduce the incidence of dropout.
Keep contact information for participants up to date.

Little et al, NEJM 2012, 367: 1355-1360;

Supported by NINDS R25NS088248 U01NS077352

Terminology: Mechanisms

- Missing Completely at Random (MCAR): missingness does not depend on either observed or unobserved variables (baseline covariates, outcome measures either observed or unobserved)
  - Coordinator forgot to collect data
- Missing at Random (MAR): missingness is independent of unobserved data, conditional on observed data
  - Patient withdrawal related to previous response
  - Missed visits related to treatment assignment
- Missing Not at Random (MNAR): missingness depends on unobserved data
  - Patient withdrawal related to non-response to treatment
  - Lower incomes may be less likely to report income

Supported by NINDS R25NS088248 U01NS077352

MCAR

- missingness does not depend on either observed or unobserved variables
- Implications for analysis
  - Complete sample is a random subsample of the full
  - Standard analysis methods valid
  - Parameter estimates unbiased
  - Power may be lost
- Evaluate assumption by comparing distributions of observed variables between dropouts and completers; significance difference implies MCAR is not valid assumption

Supported by NINDS R25NS088248 U01NS077352

MAR

- missingness is independent of unobserved data, conditional on observed data
- Implications for analysis
  - Missingness is explained by observed data
  - Parameter estimates valid as long as variables related to missingness are controlled for in analysis

Supported by NINDS R25NS088248 U01NS077352

MNAR

- missingness depends on unobserved data
- Implications for analysis
  - Missingness is NON-ignorable
  - Mechanism must be modeled

Supported by NINDS R25NS088248 U01NS077352

Analysis Strategies

- Listwise deletion (aka complete case)
  - Deletes entire record for which any of the specified variables is missing
  - Default in most statistical software packages
  - May be acceptable if minimal missingness (less than 5%?)
  - Ignores differences between completers and dropouts
  - Biased estimates if data are not MCAR
  - Loss of power
  - Does not adhere to ITT principle

Supported by NINDS R25NS088248 U01NS077352

Analysis Strategies: Single Imputation

- Substitutes an estimated value in place of the missing value
- One-time substitution
  - Last Observation Carried Forward: substitutes last observed outcome (has been criticized in the literature)
  - Worst case: substitutes worst observed value
  - Best/Worst Case: substitutes best observed value in placebo group and worst observed value in tx group
  - Mean: substitutes mean of non-missing values
  - Hot-deck: substitutes mean in homogeneous strata
  - Regression: substitutes prediction from a regression model

Supported by NINDS R25NS088248 U01NS077352

Analysis Strategies: Single Imputation

- Easy to implement
- May cause systematic bias in parameter estimate
- Does not account for uncertainty in imputation
- Increased Type I error rate

Supported by NINDS R25NS088248 U01NS077352

Analysis Strategies: Multiple Imputation

- Approach
  - Replace each missing value with a plausible value, 'drawn' from a distribution derived from an appropriate statistical model.
  - Repeat *m* times, to obtain *m* 'complete' data sets
  - Analyze each using standard procedures
  - Analysis results from each of *m* complete data sets are then combined in order to obtain a single inference
- Introduces appropriate random error into imputation process
- Yields approximately unbiased parameter estimates
- Uses all available data

Supported by NINDS R25NS088248 U01NS077352

Sample Size Calculations

- Anticipated loss should be accounted for
- Simple inflation for anticipated loss is not sufficient
- Inflation should account for attenuation of treatment effect

Supported by NINDS R25NS088248 U01NS077352