Missing Data

C. Michael Gibson, M.S., M.D.

Harvard Medical School

Chairman, PERFUSE Study Group

Founder and Chairman, WikiDoc & WikiPatient, The World's Open Source Textbook of Medicine Viewed 896 Million Times A Year

What is Missing Data?

- Missing data: Defined as absence of data for a given variable such as the outcome of a patient in a trial.

- Missing data is becoming increasingly frequent in clinical research as patients exercise their right to withdraw consent for participation in a trial consistent with the Declaration of Helsinki.

- Audience Poll:
  - If a patient discontinues study drug, does that mean they have withdrawn consent from participating in the trial?
  - If a patient discontinues study drug (or never takes it), does that mean they should no longer be followed in the trial?

Reasons for Missing Data

- Withdrawal of consent from follow-up
- Discontinuation of the study drug and failure to return for follow-up (should not happen)
- Lost to follow-up (should not happen)
- Outcome cannot be assessed for other reasons (i.e., due to incarceration)
- Individual missed visits can lead to missing outcome data at selected time points

Withdrawal of Consent vs Drug DC vs Lost to F/U

	Study Drug DC	Withdrew Consent	Clinical Follow-Up Available
Withdrew consent	Yes	Yes	No
Discontinued Study Drug	Yes	No	Yes
Lost to Follow-Up	?	No	No, but could be

C. Michael Gibson, M.S., M.D.

Levels of Follow-up in a Clinical Research Study

1. Pt takes the investigational product (IP) and attends all visits according to protocol – ideal scenario
2. Pt does not take IP for whatever reason, but agrees to attend all clinic visits according to protocol
3. Pt does not attend in person, but agrees that the site calls him by telephone
4. Pt refuses tel contact, but agrees that his data be obtained from a third party (eg GP) or from e-records without bothering him
5. Data regarding vital status from public records

Importance of Missing Data

- Missing data may be associated with an inability to tolerate study drug.
- By only analyzing patients with non-missing data, you exclude patients who were unable to tolerate study drug and the analysis then applies only to patients who successfully tolerated the study drug
- We obviously prescribe drugs to all patients with a disease, and we cannot identify those patients who will or will not tolerate the drug prospectively.

Regulatory Hypothesis

- A side effect like bleeding causes patients to either stop study drug or withdraw consent
- Patients who bleed go on to die after they exit from the study
- Since there is no follow-up of these patients who bled (the data is missing) the adverse consequences of bleeding or other side effect are not captured
- Despite favorable outcomes in the patients who did not bleed or have a side effect, these favorable outcomes may be offset by the adverse events in the patients who have missing data
- Regulators may therefore assume that all study drug patients with missing data died, and that all control patients with missing data lived

Missing Data & ITT Analyses

- ITT Analysis: All the patients for all of the trial. Once randomized, analyzed.
- Missing data erodes the ability to perform an intention to treat (ITT) analysis.
- An ITT analysis includes all patients randomized to therapy irrespective of protocol deviations, study drug discontinuation, or patient withdrawal.
- An ITT analysis maintains randomization and minimizes the potential for bias.
- **If you do not follow all patients for the entire duration of trial due to missing data, you cannot perform an ITT analysis.**
- You are actually performing a modified ITT analysis that censors follow-up at the time of last contact with the patient.

The Hierarchy of Missing Data

- Missing Completely at Random
  - Least introduction of bias
- Missing at Random
- Missing Not at Random
  - Greatest potential for introduction of bias.

Missing Completely at Random (MCAR)

- Missing Completely at Random data is independent of observed and non-observed data, and is therefore not related to the independent variables or the outcome.

Examples:

- Loss of study files
- Equipment malfunction
- Data entered incorrectly
- Weather conditions that lead to a missed patient visit.

Missing at Random (MAR)

- Missing at Random data is related to independent variables (i.e., age, race, gender), but is not related to the outcome.

Examples:

- Elderly patients drop out from an intervention due to their physical condition, but this variable is not related to the outcome.
- Patients drop out because of known baseline characteristics.

Missing Not At Random (MNAR)

- Missing Not at Random data is related to the outcome.
  - This is the worst type of missing data because dropouts are related to the therapy or intervention under investigation.
  - The pattern of missing data is related to unobserved data, making it impossible to use other values from the dataset to predict missing values.

Examples:

- Patients discontinue study drug and withdraw from trial because they cannot tolerate a side effect of the drug like bleeding.
- Patients miss a visit because they had an outcome.

Preventing Missing Data

Educate the patients about the importance of follow-up, even if study drug is discontinued early

- During informed consent
  - Informed Consent Form should distinguish between withdrawal of consent and early discontinuation of study drug.
  - The patient should prospectively be made aware that if they discontinue study drug, they must still be followed - up
- At every patient visit
  - Reinforce need for complete follow-up at visits

Preventing Missing Data

Limit individual missed visits

- Keep patients engaged
  - Study newsletters, visit reminders, check-up calls
- Resolve barriers to visit attendance
  - Reimburse parking and travel costs
  - Flexible follow-up schedule
  - Alternate methods of follow-up (i.e., Skype visits instead of in-person visits if allowed by IRB and study)

Preventing Missing Data

- If a patient discontinues study drug and the PI classifies them as having withdrawn consent, the PI must write a letter to the Executive Committee of the trial justifying lack of follow-up

Limitations in the *Late* Collection of Missing Data

- Do not wait until the end of the study or after the study is over to solve problems with missing data
- When contacting patients after a long period of time has elapsed since the scheduled follow-up visit, beware of recall bias:
  - Patient may not remember side effects, bleeding episodes months after the follow-up visit was to occur, and may not remember dates of events
- Therefore, re-establish contact with patients as early as possible
- Obtain medical records to verify patient narratives

Process for Approaching Patients Who Have Withdrawn Consent

If permitted by national and local laws, seek permission of IRB to re-consent patients who have withdrawn consent to collect data regarding the study outcome and vital status

Process for Locating Patients Who Are Lost to Follow-up

If permitted by national and local laws, a patient locator service can be used to locate patients lost to follow-up

If permitted by national and local laws, a death registry or other public records can be used to identify patients who have died

eCRF Should be Designed to Classify Missing Data

The eCRF must capture which of the 4 following categories the patient with missing data belongs to

- Withdrawal of consent from follow-up
- Discontinuation of the study drug and failure to return for follow-up (should not happen)
- Loss to follow-up
- Other: Outcome cannot be assessed for other reasons (i.e., due to incarceration)

Consequences of High Rates of Missing Data in a Trial

- Regulators may require you to use patient locator service / death registries to track down all patients
- Even if you ascertain patient's *vital status* using death registry, you will *not* ascertain *primary endpoint*
- Mortality will likely be the same after withdrawal of consent, which will be interpreted as "the drug did not reduce mortality"
- You will need to re-analyze mortality data using newly ascertained vital status in missing patients to show trial was still significant

Strategies to Deal with Missing Data

- You will need to calculate not only number of missing patients, but missing time of follow-up:
  - A patient who withdraws from the trial the day before it ends counts as a missing patient, but only 0.3% of the patient's exposure time was missing in a 1 year trial.
- You will need to adjust for the duration of the trial to compare the amount of "missingness" to other trials.
  - A trial that goes on for two years may have twice as much missing data as a trial that goes on for one year
  - In an event driven trial, there will be varying durations of exposure (e.g. months to years), and "missingness" cannot be compared to a fixed duration trial

Quantitative Example: 3 scenarios

Scenario	Event rate in placebo group	Event rate in active treatment group	RRR	P-value (z-test for proportions)
1 No discontinuation; No loss to follow-up	20%	15%	25%	P=0.004
2 Discontinuation without loss to follow-up	20% (observed event rate 20% among adherent subjects, 20% among non-adherent subjects	16% (observed event rate 15% among adherent subjects, 20% among non-adherent subjects	20%	P=0.02
3 Loss to follow-up in both groups, worst case analysis	19.6% (observed event rate 20% in 98% of subjects, assumed event rate 0% in 2% of subjects)	16.7% (observed event rate 15% in 98% of subjects, assumed event rate 100% in 2% of subjects)	14.8%	P=0.10

Conclusions

- Discontinuations of treatment should be avoided/reversed whenever possible
- Loss to follow-up has a profound impact on validity and statistical interpretation of a trial
- A significant number of patients lost to follow-up can have negate the findings in a clinical trial, even when the test agent is highly effective.
- Actions to prevent loss to follow-up begin during recruitment and continue until the trial ends