Biostatistics Workshop 2008

$\sim$ Longitudinal Data Analysis $\sim$

Session 7

GARRETT FITZMAURICE

Harvard University

Unequal n_i per Subject or Unbalanced Designs

Basically, all the methods that we have discussed (ML, REML, GLS/GEE) can handle situations in which $n_i \neq n$ for all i (i.e., unequal number of observations per subject).

That is, modern regression methods can handle unbalanced longitudinal designs with relative ease.

However, need to be more careful when $n_i \neq n$ for all i due to missingness.

For most designed studies, we *plan* on measuring the same number of outcomes, so if $n_i \neq n$ for all i, then some outcomes are *missing*.

Missing Data

In longitudinal studies missing data are the rule not the exception.

The term "missing data" is used to indicate that an intended measurement could not be obtained.

With missing data there must necessarily be some loss of information.

Of greater concern, missing data can introduce bias and result in misleading inferences about change over time.

Let Y denote the complete response vector which can be partitioned into two sub-vectors:

(i) Y^O the measurements observed

(ii) Y^M the measurements that are missing

If there were no missing data, we would have observed the complete response vector Y.

Instead, we get to observe Y^O.

The main problem with missing data is that distribution of the observed data may not be the same as distribution of the complete data.

When data are missing we must carefully consider the reasons for missingness.

Estimation of β with missing data depends on the missing data mechanism.

The missing data mechanism is a probability model for missingness:

- Missing Completely at Random (MCAR)
- Missing at Random (MAR)
- Not Missing at Random (NMAR)

Missing Completely at Random (MCAR)

MCAR: probability that responses are missing is unrelated to either the specific values that, in principle, should have been obtained (the *missing* responses) or the set of *observed* responses.

MCAR: probability responses are missing is independent of Y^O and Y^M.

Missingness is simply the result of a chance mechanism that is unrelated to either observed or unobserved compon ents of the outcome vector.

Consequently, observed data can by thought of as a random sample of the complete data.

Features of MCAR

The means, variances, and covariances are preserved.

- Can use ML/REML estimators for β
- More generally, we can use GLS/GEE estimator with any "working" assumption for the covariance; distributional assumption for Y_{ij} is not necessary
- If we use GLS/GEE estimator with incorrect "working" assumption for the covariance, then must use "empirical" or "sandwich" variance estimator for $\text{Cov}(\widehat{\beta})$

$\Longrightarrow$ With complete data or data MCAR, distributional assumption is not required.

Missing at Random (MAR)

MAR: probability that responses are missing depends on Y^O, but is conditionally independent of Y^M.

Note 1: If subjects are stratified on the basis of similar values for the responses that have been observed, then within strata missingness is simply the result of a chance mechanism unrelated to unobserved responses.

Note 2: Because missingness depends on observed responses, the distribution of Y_i in each of the distinct strata defined by the patterns of missingness is not the same as the distribution of Y_i in the target population.

The "completers" are a biased sample from the target population.

Features of MAR

Means, variances, and covariances are not preserved.

This implies that sample means, variances, and covariances based on either the "completers" or the available data are biased estimates of the corresponding parameters in the target population.

In general, standard WLS/GEE approach may lead to inconsistent inferences.

Alternatively, a "weighted GEE" approach can be used, incorporating estimated response probabilities.

However, the likelihood is preserved.

This implies that ML estimation of β is valid when data are MAR provided the joint distribution has been correctly specified.

For multivariate normal, this requires correct specification of not only the model for the mean response, but also the model for the covariance among the responses.

In a sense, ML estimation allows the missing values to be validly "predicted" or "imputed" using the observed data and a correct model for the joint distribution of the responses.

Not Missing at Random (NMAR)

NMAR: probability that responses are missing is related to the specific values that should have been obtained.

An NMAR mechanism is often referred to as "non-ignorable" missingness.

Challenging problem and requires modelling of missing data mechanism; moreover, specific model chosen can drive results of analysis.

Sensitivity analyses is recommended.

Dropout

Longitudinal studies often suffer from problem of attrition; i.e., some individuals "drop out" of the study prematurely.

Term *dropout* refers to special case where if Y_{ik} is missing, then Y_{ik+1}, ..., Y_{in} are also missing.

This gives rise to so-called "monotone" missing data pattern displayed in Figure 1.

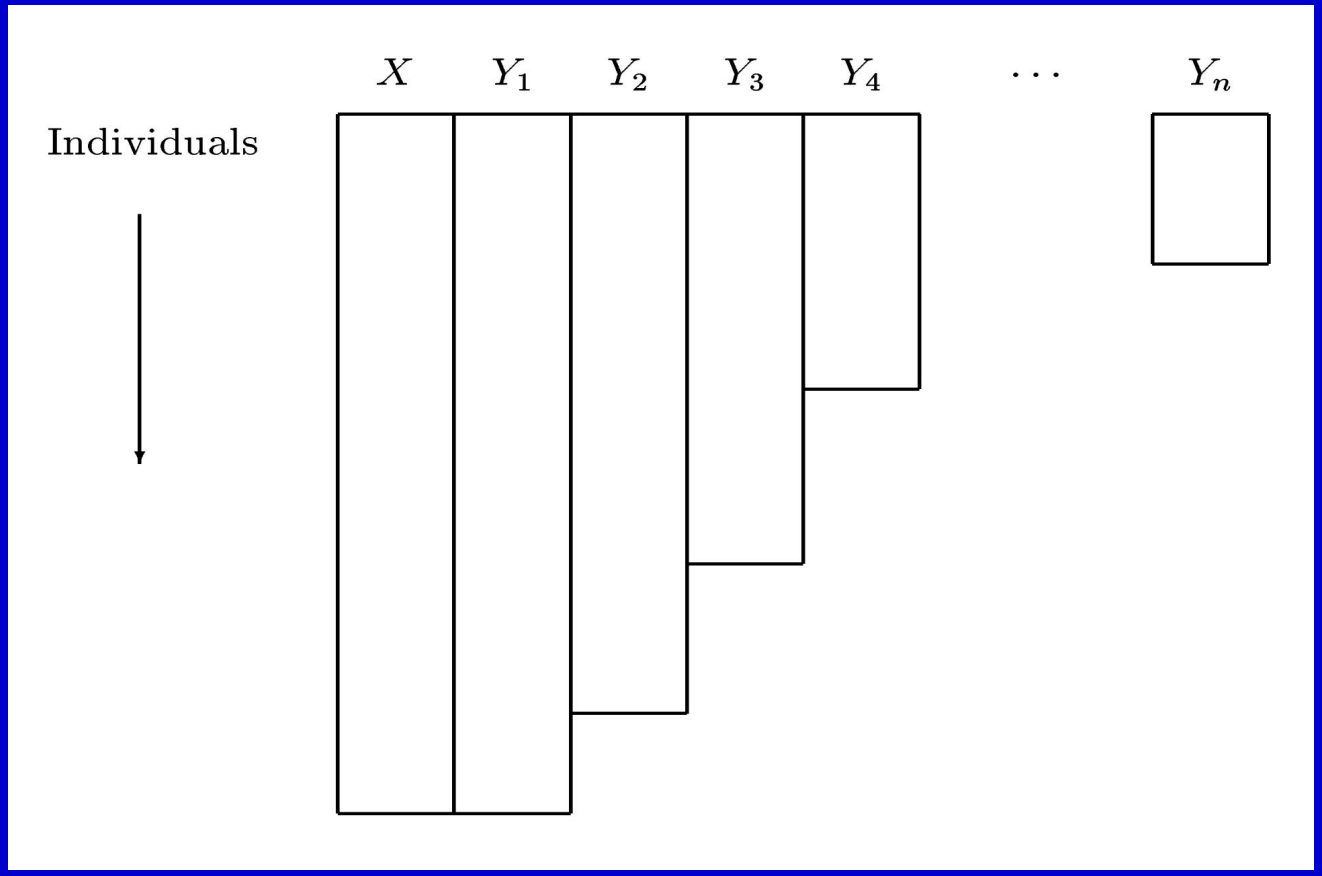

Figure 1: Schematic representation of a monotone missing data pattern for dropout, with Y_j more observed than Y_{j+1} for $j = 1, ..., n-1$.

When there is dropout, key issue is whether those who "drop out" and those who remain in the study differ in any further relevant way.

If they do differ, then there is potential for bias.

The taxonomy of missing data mechanisms (MCAR, MAR, and NMAR) discussed earlier can be applied to dropout.

Common Approaches for Handling Dropout

Complete-Case Analysis:

Exclude all data from the analysis on any subject who drops out.

That is, a so-called "complete-case" analysis can be performed by excluding any subjects that do not have data at all intended measurement occasions.

This method is very problematic and is rarely an acceptable approach to the analysis.

It will yield unbiased estimates of mean response trends only when dropout is MCAR.

Even when MCAR assumption is tenable, complete-case analysis can be immensely inefficient.

Available-Data Analysis:

General term that refers to a wide collection of techniques that can readily incorporate vectors of repeated measures of unequal length in the analysis.

Standard applications of WLS/GEE are available-data methods.

In general, available-data methods are more efficient than complete-case methods.

However, many available-data methods yield biased estimates of mean response trends unless dropout is MCAR.

Imputation

Imputation: substitute or fill-in the values that were not recorded with imputed values.

Once a filled-in data set has been constructed, standard methods for complete data can be applied.

Validity of method depends on how imputation is done.

Methods that rely on just a single imputation fail to acknowledge the uncertainty inherent in the imputation of the unobserved responses.

"Multiple imputation" circumvents this difficulty.

Multiple Imputation (MI): Missing values are replaced by a set of m plausible values, thereby acknowledging uncertainty about what values to impute.

Typically, a small number of imputations, for instance, $5 \leq m \leq 10$, is sufficient.

The m filled-in data sets produce m different sets of parameter estimates and their standard errors.

These are then combined to provide a single estimate of the parameters of interest, together with standard errors that reflect the uncertainty inherent in the imputation.

"Last Value Carried Forward" (LVCF):

One widely used imputation method, especially in clinical trials, is LVCF.

Regulatory agencies such as FDA seem to encourage the continuing use of LVCF.

LVCF makes a strong, and often very unrealistic, assumption that the responses following dropout remain constant at the last observed value prior to dropout.

There appears to be some statistical folklore that LVCF yields a *conservative* estimate of the comparison of an active treatment versus the control.

This is a gross misconception!

Except in very rare cases, we do not recommend the use of LVCF as a method for handling dropout.

Variations on the LVCF theme include baseline value carried forward and worst value carried forward.

Imputation methods based on drawing values of missing responses from the conditional distribution of the missing responses given the observed responses have a much firmer theoretical foundation.

Then subsequent analyses of the observed and imputed data are valid when dropouts are MAR (or MCAR).

Furthermore, multiple imputation ensures that the uncertainty is properly accounted for.

Model-Based Imputation:

There is a related form of "imputation" where missing responses are *implicitly* imputed by modelling joint distribution of Y_i, $f(Y_i|X_i)$.

When dropout is MCAR or MAR, likelihood-based methods can be used based solely on the marginal distribution of the observed data.

In a certain sense, the missing values are validly predicted by the observed data via the model for the conditional mean of the missing responses given the observed responses (and covariates).

However, likelihood-based approaches require model for $f(Y_i|X_i)$ must be correctly specified (e.g., any misspecification of the covariance will, in general, yield biased estimates of the mean response trend).

Weighting Methods

In weighting methods, under-representation of certain response profiles in the observed data is taken into account and corrected.

These approaches are often called "propensity weighted" or "inverse probability weighted" methods.

Basic Idea: Base estimation on the observed responses but weight them to account for the probability of remaining in the study.

Intuition: Each subject's contribution to the weighted analysis is replicated to count for herself and for those subjects with the same history of responses and covariates, but who dropout.

Propensities for dropout can be estimated as a function of observed responses prior to dropout and covariates.

For example, GEE approach can be adapted to handle data that are MAR by making adjustments to the analysis for the propensities for dropout.

Inverse probability weighted methods were first proposed in sample survey literature, where the weights are known.

In contrast, with dropout the weights are not known, but must be estimated from the observed data.

In general, weighting methods are valid provided model that produces the estimated weights is correctly specified.

Finally, propensity weights can be used for imputation (see later).

Case Study: Clinical Trial of Contracepting Women

Randomized clinical trial comparing two doses of a contraceptive:
4 injections of 100 mg or 150 mg of DMPA, given at 90-day intervals.

Woman completed a menstrual diary that recorded any vaginal bleeding pattern disturbances.

Outcome of interest is a repeated binary response indicating whether or not a woman experienced amenorrhea (absence of menstrual bleeding) during follow-up intervals.

A total of 1151 women completed the menstrual diaries.

Dropout: There was substantial dropout for reasons that were thought likely to be related to the outcome.

More than one third of the women dropped out of the trial:

- 17% dropped out after 1st 90 day interval
- 13% dropped out after 2nd 90 day interval
- 7% dropped out after 3rd 90 day interval

When the dropout rates are broken down by dose group, the rates were marginally higher in the 150 mg dose group.

Analytic Goal: Estimate dosage specific rates of amenorrhea that would have been observed in the absence of dropout.

Let $Y_{ij} = 1$ if i^{th} woman experienced amenorrhea in the j^{th} injection interval.

We considered following logistic regression model for marginal mean:

$$\text{logit}(\mu_{ij}) = \beta_0 + \beta_1 \mathtt{t_{ij}} + \beta_2 \mathtt{t_{ij}^2} + \beta_3 \mathtt{dose_i} + \beta_4 (\mathtt{t_{ij}} \times \mathtt{dose_i}) + \beta_5 (\mathtt{t_{ij}^2} \times \mathtt{dose_i}),$$

where $\mu_{ij} = \Pr(Y_{ij} = 1)$, $\mathtt{t} = 0, 1, 2, 3$ for the four consecutive 3-month injection intervals, $\mathtt{dose} = 1$ if randomized to 150 mg of DMPA, and $\mathtt{dose} = 0$ otherwise.

To account for within-subject association among repeated measures, we fit six separate pairwise log odds ratios.

We considered following methods for handling dropout:

1. Complete-Case (CC)

2. Available-Data (AD)

3. LVCF imputation (for illustrative purposes)

4. Multiple Imputation (MI) based on Propensity Scores

Multiple Imputation based on Propensity Scores

Propensity score methods require a model for the probability of dropout.

Within each dose group, we consider a sequence of logistic regression models that assume log odds of dropout depends on all past observed responses.

For example, model for dropout at k^{th} occasion is given by

$$\log\left\{\frac{\Pr(D_i = k | D_i \geq k, Y_{i1}, ..., Y_{ik-1})}{\Pr(D_i > k | D_i \geq k, Y_{i1}, ..., Y_{ik-1})}\right\} = \theta_1 + \theta_2\, Y_{i1} + \cdots + \theta_k\, Y_{ik-1}.$$

Based on the estimated parameters from this model, a propensity score is obtained for each observation at each occasion of dropout.

Following procedure used at each dropout occasion to draw imputed values:

(i) Observations are sorted into eight groups based on the propensity scores.

(ii) Within each group, let N^O denote number of individuals with observed values for the response at that occasion, and N^M denote number of individuals with missing values. We randomly select N^O observations with replacement from the observed values for the response.

(iii) Finally, the N^M values for the missing responses are then randomly selected with replacement from the random sample of observed values drawn in (ii).

This three-step procedure is repeated sequentially at each occasion to fill in all of the missing values.

To create 10 imputations, steps (ii) and (iii) are repeated 10 times.

Table 1: Estimated marginal rates of amenorrhea for quadratic trend model using GEE under four different methods for handling dropouts: complete-case (CC), last value carried forward (LVCF), available-data (AD), and multiple imputation (MI).

Method	Time	100 mg	150 mg	Difference	SE	Z
CC	12 months	0.502	0.540	0.038	0.037	1.03
LVCF	12 months	0.437	0.498	0.061	0.029	2.10
AD	12 months	0.517	0.572	0.055	0.036	1.52
MI	12 months	0.517	0.572	0.056	0.034	1.64

Results

Differences in results are more easily discerned by considering dose-specific estimated rates of amenorrhea at month 12 (see Table 1).

Results of complete-case, available-data, and MI analyses suggest no treatment difference.

GEE analysis based on LVCF imputation produces lower estimated rates of amenorrhea.

GEE analysis based on LVCF imputation suggests that there are treatment differences in the estimated rates of amenorrhea at the end of the trial.

Result is driven by smaller SE (failing to account for uncertainty).

Results from analysis based on multiple imputation are very similar to those obtained from the available-data analysis.

Both the point estimates of the rates of amenorrhea and their standard errors are similar.

Note, this similarity cannot be expected in general.

Summary

In longitudinal studies missing data are the rule not the exception.

Missing data have two important implications:
(i) loss of information, and
(ii) validity of analysis.

The loss of information is directly related to the amount of missing data; it will lead to reduced precision (e.g., larger SEs, wider CIs) and reduced statistical power (e.g., larger p-values).

The validity of the analysis depends on assumptions about the missing data mechanism.

FURTHER READING

Diggle, P.J., Heagerty, P., Liang, K-Y. and Zeger, S.L. (2002). *Analysis of Longitudinal Data* (2nd ed.). Oxford University Press. (See Chapter 13).

Fitzmaurice, G.M., Laird, N.M. and Ware, J.H. (2004). *Applied Longitudinal Analysis*. Wiley. (See Chapter 14).