

A Review of Methods for Handling Nonignorably Missing Data

Rachel Tsong and Rebekah Yu

11/25/2019

Plan of Presentation

- I. Give background information and motivation for our topic
- II. Describe Non-Bayesian methods of dealing with nonignorable missing data
- III. Describe newly proposed method for modeling nonignorable missing data with a Bayesian framework
- IV. Compare advantages and disadvantages of Bayesian Methods for nonignorable missing data

Background and Motivation

Missing data are often a concern in longitudinal clinical trials. It is not uncommon for subjects enrolled in longitudinal trials to have intermittent missingness due to missed visits or permanent missingness due to drop-out. These missing data can be classified into three categories: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Data are considered to be MCAR if the missingness is independent of both observed and unobserved data. Data are MAR if the missingness is independent of the unobserved data conditioned on the observed data. MCAR and MAR are classified as “ignorable” mechanisms of missingness because the missing data mechanism does not need to be specified in the log-likelihood. However, MNAR data are when the probability of missingness depends on some unobserved variable, and this missing data mechanism is referred to as “nonignorable” because the missingness mechanism needs to be incorporated into the likelihood estimates. Inappropriate handling of missing data can lead to biased inferences, thus it is of great clinical significance to develop methods for handling missing data.

Bayesian methods are a useful approach for dealing with missing data because they consider missing data as random variables whose posterior distributions can be obtained by specifying priors and missing covariate distributions. Bayesian approaches have been shown to produce more reliable results compared to traditional missing data models such as multiple imputation. We reviewed Bayesian models for nonignorably missing

data, including an in-depth review of a method for longitudinal binary responses developed by Wu et al. (2018).

Non-Bayesian Methods

The following models are popularized models for dealing with nonignorable missing data. Each will be described briefly.

1. Selection Model
2. Pattern-Mixture Model
3. Share-parameter model

Newly Proposed Model (Bayesian Inference)

We will describe the Bayesian approach for modeling missing data. This approach regards missing data as a random variable and constructs informative priors from observed data. In addition, Wu et al. (2018) proposes a new method for defining priors to establish a more proper posterior distribution. The Bayesian framework can be used to compare the performance (DIC) of fitting a nonignorable model versus an ignorable model to data where missingness is present. Results from simulated data comparing the two models will be presented.

Advantages and Disadvantages

Some papers propose that the Bayesian approach for dealing with nonignorable missing data achieves more reliable results. However, Bayesian approaches are computationally intensive and requires correct specification of prior distributions which may be difficult or impossible in some cases.

References

- Ma Z, Chen G. Bayesian methods for dealing with missing data problems. *Journal of the Korean Statistical Society*. 2018;47(3):297-313. doi:10.1016/j.jkss.2018.03.002.
- Wu J, Ibrahim JG, Chen MH, Schifano ED, Fisher JD. Bayesian Modeling and Inference for Nonignorably Missing Longitudinal Binary Response Data with Applications to HIV Prevention Trials. *Stat Sin*. 2018;28:1929–1963. doi:10.5705/ss.202016.0319