

## Chapter 9

# Nonignorable Nonresponse

The techniques discussed in the preceding sections all can accommodate missing data and/or unbalanced designs, but valid inferences require assumptions on the missing data mechanism (MDM). The GEE methodology requires missing completely at random (MCAR); the likelihood approaches are valid with missing at random (MAR) nonresponse mechanisms, but then further require that the likelihood be correctly specified. We note that the GEE can be adapted to handle MAR mechanisms as well by using nonresponse weights, similar to nonresponse adjustments used in sample surveys (Robins, Rotnitzky and Zhao, 1995).

If the MDM depends upon the unobserved responses, given the observed responses, it is said to be nonignorable. Then the inferential issues are more complex, because here we need to make unverifiable assumptions on the MDM in order to have valid inferences. We consider some examples, many of which are univariate since the issues are not fundamentally different for the multivariate case.

**Example 1.** In estimating mean income (or a regression of income on covariates), nonresponse is often assumed to be more prevalent among those with very high or very low incomes (Greenlees, *et al.*, 1982).

**Example 2.** Studies of factors which influence wages in workers may be biased by the omission of workers who are not currently in the workforce due to unemployment (Heckman, 1976). Heckman used the term selection bias to describe such self-selected samples and noted the connection to nonignorable nonresponse. He proposed methods for dealing with nonresponse in the univariate setting based on estimating the parameters of a model for  $P(\text{missing})$ . Such approaches are generally referred to as selection model approaches.