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.

Example 3. Many voters respond "don't know," "undecided," or "will not vote" in pre-election surveys. Such respondents are usually ignored and predictions are based on respondents who give a preference. However, voter unwillingness to commit to an answer may be nondifferential depending upon who they ultimately vote for. Baker and Laird (1988) reanalyze the 1948 Truman–Dewey preelection polls which were notorious for failing to predict a win for Truman. Their analysis suggests that voters in the "don't know" or "undecided" categories were far more likely to favor Truman; reanalysis under this scenario predicts an outcome very close to the election and to the voter exit polls.

We now turn to some longitudinal data examples.

Example 4. In cancer chemotherapy trials, increasing emphasis is placed on evaluating quality-of-life of patients on different therapies. Quality-of-life is typically measured quarterly with a self-report questionnaire. The questionnaire is long and there is considerable missing data even among those patients who remain on the study protocol. Cancer chemotherapy patients often have periods of debilitating pain or illness, and a plausible assumption is that these patients would be less likely to make the effort to respond under these circumstances. Nonresponse at a particular occasion is less likely to be predicted by quality-of-life score at the previous occasion (which might be observed), but more likely depends on current unobserved score.

Other examples of intermittent nonignorable response in longitudinal surveys are more difficult to find. A more common problem in this setting is monotone missingness, or dropouts, when subjects are removed from study, dropout or otherwise become unavailable for study at some point. If $P(\text{dropout at } t)$ depends only on observed covariates and past history, i.e., y_{ij}'s observed prior to dropout, then dropout is ignorable. If $P(\text{dropout at } t)$ depends upon the unobserved response which would have been obtained (say Y_{it}) had the subject not dropped out, dropout is nonignorable. Several examples can be found in Diggle and Kenward (1994).

Example 5. The U.S. government uses an annual panel survey of doctoral level scientists and engineers to estimate current unemployment rates among scientists and engineers. A follow-up of those ceasing to participate in the survey revealed that a disproportionate number had moved abroad, relative to those remaining in the survey. Since the government is only interested in employment in the U.S., this is a clear example of nonignorable dropout.

Example 6. In many clinical trials, patients can be removed from the protocol for reasons related to the outcome of interest. For example, in a drug trial of psychosis in schizophrenics, the endpoint was reduction in psychosis after six weeks on drug (Hogan and Laird, 1997a). But many patients were removed from study drug early due to lack of effect. In cases such as this, one might argue that dropout is ignorable since it can be predicted by the observed prior outcomes, although the indications used by clinicians to remove patients from a study protocol may not be quantified in the data available for analysis.

9.1 Terminology

Let us assume that for each subject, the vector of complete responses, Y_i can be partitioned into $Y_i^{\text{OBS}} = (Y_{i1}, \ldots, Y_{ik})^T$ and $Y_i^{\text{MIS}} = (Y_{ik+1}, \ldots, Y_{in})^T$, and d_i denotes a subject's dropout time where $t_k < d_i \leq t_{k+1}$. This simple model implies Y_{ij} is always observed prior to d_i and never after. Of course, if $d_i > t_n$, subjects do not drop out.

Nonignorable Nonresponse:

If

$$P(t_k < d_i \leq t_{k+1} \mid Y_i, \psi) = P\left(t_t < d_i \leq t_{k+1} \mid Y_i^{\text{OBS}}, Y_i^{\text{MIS}}, \psi\right), \quad (9.1)$$

i.e., the probability of nonresponse due to dropout depends on future unobserved values Y_i^{MIS}, even conditional on the past, Y_i^{OBS}, then dropout is nonignorable. The term nonignorable comes from the fact that valid likelihood based inferences require specification of the nonresponse mechanism (which is also called the selection model), and maximizing the observed data likelihood of (Y_i^{OBS}, d_i). This will be described in Section 9.2.

The term informative dropouts is often used in this setting. We define **non**-informative dropout to mean:

Noninformative Dropout:

$$f\left(Y_i^{\text{MIS}} \mid X_i, Y_i^{\text{OBS}}, d_i\right) = f\left(Y_i^{\text{MIS}} \mid X_i, Y_i^{\text{OBS}}\right) \quad (9.2)$$

where the same partitioning of $Y_i^{\text{MIS}}, Y_i^{\text{OBS}}$ is used in both sides of (9.2), and determined by $t_k < d_i \leq t_{k+1}$. Noninformative thus implies that the **fact** of dropout implies no new information about future values that is not already contained in a patient's history, (X_i, Y_i^{OBS}). Conversely, if equality does **not** hold in (9.2), then dropout is said to be informative, i.e., the fact of dropping out is informative about future (missing) values, even conditional on past history.

It is straightforward to see that ignorable dropout and noninformative are equivalent concepts, and informative and nonignorable are as well. An advantage of the informative (noninformative) representation is that it makes it clear what is inherently inestimable in this setting, and what features of the model the results will be sensitive to. The basic problem is that without additional assumptions, there are no data on those observed after d_i to estimate responses for an individual with dropout d_i.

9.2 Methodology: General Comments

There are two general model-based approaches to handling nonignorable nonresponse (Hogan and Laird, 1997b): selection-modeling and mixture-modeling. In this section we give a brief overview of each; detailed case studies are given in Section 9.3.

Selection Modeling. As noted earlier, the term selection modeling refers to approaches which specify both the distribution of Y_i and R_i (or d_i) and base inference on the joint distribution. The contribution of the ith subject to the likelihood can be expressed as

$$L_i(\beta, \theta, \psi) = \int f(Y_i \mid \beta, \theta)\; p(d_i \mid Y_i, \psi) dY_i^{\mathrm{MIS}}.$$

Since $p(d_i \mid Y_i, \psi)$ depends upon the unobserved Y_i's, estimation of the model parameters is at best difficult, and in general sensitive to model specification. To see why, consider the simple univariate case where we assume that $Y_i \sim N(\beta, \theta)$ and p (nonresponse) is monotone in Y_i. Then intuitively, if nonresponse is independent of Y_i, our sample histogram using the observed data should look approximately normal; if P(nonresponse) is monotone increasing in Y_i, it will be skewed the left, and right skewed if P(nonresponse) is decreasing in Y_i:

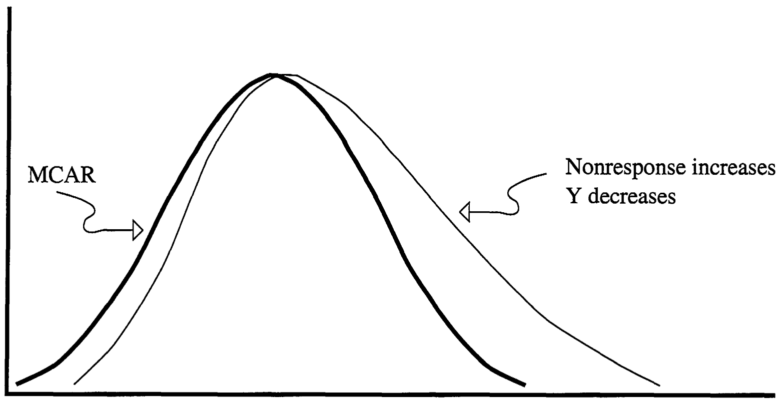

FIGURE 9.1.

Thus estimation of P(nonresponse) relies heavily on assumed normality. If we assume log-normality for Y_i, we could see quite different results.

In the case with covariates, multivariate responses and arbitrary patterns of missingness, it is generally not possible to permit all possible models because of nonidentifiability problems. In fact, nonidentifiability can arise in the simple response setting where $n = 2$, each Y_{ij} is dichotomous, and there is missingness in Y_{i2} only. In this case, one must either impose restrictions on $f(Y_i)$ or on $f(R_i \mid Y_i)$ where R_i is an indicator of missing Y_{i2}, to estimate the model parameters. Little and Rubin (1989, Chapter 11) show that the following models are **not** identifiable:

(1) $\quad f(Y_i) = f_1(Y_{i1})f_2(Y_{i2}) \quad \text{and} \quad f(R_i \mid Y_i) = f(R_i \mid Y_{i2}),$

(2) $\quad f(Y_i, R_i) = f_1(Y_{i1} \mid R_i)f_2(Y_{i2} \mid R_i)f(R_i),$

(3) $\quad f(Y_i, R_i)$ completely general,

and, in addition, a model which specifies that all two-way associations are non zero: $(Y_{i1}, R_i), (Y_{i2}, R_i)$ and (Y_{i1}, Y_{i2}). Thus the only possible nonignorable model in this setting is

(4) $\quad f(Y_i)$ arbitrary and $f(R_i \mid Y_i) = f(R_i \mid Y_{i2}).$

Applications of selection modeling in this setting typically presume a simplified model for dropout or nonresponse, partly in order to overcome technical limitations, partly to ensure identifiability and partly to have reasonable assumptions about the MDM. In the dropout (monotone missing) setting a typical assumption is P(dropout) depends only on current unobserved value. Diggle and Kenward (1994) and Troxel *et al.* (1998) assume missingness depends only on current unobserved value.

Random Effects Selection Models

A somewhat different approach, which is especially suitable for longitudinal data, is random effects selection modeling. Here we allow P(dropout) to be a function of unobserved random effects. Recall the random effects model

$$Y_i = X_i\beta + Z_i b_i + e_i,$$

where as usual, we take b_i and e_i to be independent $N_q(0, D)$ and $N_n(0, \sigma^2 I)$. We now further assume that d_i has a distribution which depends upon b_i. For example, suppose we are fitting linear models, so that

$$Z_i = \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix}$$

and (b_{0i}, b_{1i}) represent the ith individual's unobserved, random, residual slope and intercept. If we assume that

$$f(d_i \mid Y_i, b_i, \psi) = f(d_i \mid b_i, \psi),$$

i.e., dropout given b_i does not depend upon Y_i, then we still have non-ignorable dropout since $f(d_i \mid Y_i^{\text{OBS}}, Y_i^{\text{MIS}}, \psi)$ will in general depend on Y_i^{MIS}.

An attractive feature of this model is that it follows immediately that

$$\underset{(n_i \times 1)}{Y_i^{\text{OBS}}} = \underset{(n_i \times p)(p \times 1)}{X_i^{\text{OBS}}\beta} + \underset{(n_i \times q)(q \times 1)}{Z_i^{\text{OBS}} b_i} + \underset{(n_i \times 1)}{e_i}$$

and the marginal distribution of (Y_i^{OBS}, d_i) is obtained as

$$f(Y_i^{\text{OBS}}, d_i) = \int_{R^q} f(Y_i^{\text{OBS}} \mid b_i, X_i, \beta, \theta)\ f(b_i \mid D)\ f(d_i \mid b_i, \psi) db_i.$$

This model was originally proposed by Wu and Carroll (1988) who assumed that the probit of the probability of dropout at each occasion, conditional on being at risk, was linear in b_{1i} and an indicator of period. The probit model allows a closed form expression for the integral, in terms of the cumulative normal, but the complex nature of the computations led them to seek alternative approximations.

Schluchter (1992) and DeGruttola and Tu (1994) independently noticed that if d_i (or some transformation of d_i) is $N_1(u_i^T\psi_0 + b_i^T\psi_1, \tau^2)$

where u_i is a vector of covariates, then the marginal distribution of $(Y_i^{\mathrm{OBS}}, d_i)$ is easily obtained as

$$\begin{pmatrix} Y_i^{\mathrm{OBS}} \\ d_i \end{pmatrix} = N\left[\begin{pmatrix} X_i^{\mathrm{OBS}}\beta \\ u_i^T\psi_0 \end{pmatrix} \begin{pmatrix} Z_i^{\mathrm{OBS}} D Z_i^{\mathrm{OBS}^T} + \sigma^2 I & Z_i^{\mathrm{OBS}} D\psi_1 \\ \psi_1^T D Z_i^{\mathrm{OBS}^T} & \psi_1^T D\psi_1 + \tau^2 \end{pmatrix}\right].$$

The marginal distribution for Y_i^{OBS} is the same as the usual random effects distribution, but d_i has a factor-analysis structure rather than a random effects structure. That is, the model for Y_i^{OBS} resembles ordinary random effects, while that for d_i resembles factor analysis. If there is no censoring of the d_i's then the analysis is relatively straightforward. In many cases, however, d_i may be censored. In the case of informative dropouts in clinical trials, d_i is time to dropout due to endpoint related reasons; hence noninformative dropouts (e.g., withdrawals due to non informative reasons) and completers will be censored.

Notice that this particular random effects selection model also implies that

$$E(Y_i^{\mathrm{OBS}} \mid X_i, d_i) = X_i^{\mathrm{OBS}}\beta + Z_i^{\mathrm{OBS}} D\psi_1(d_i - u_i^T\psi_0)/(\psi_1^T D\psi_1 + \tau^2),$$

i.e., the conditional mean of Y_i^{OBS} depends linearly upon d_i. From the joint distribution of (Y_i, b_i, d_i), it is also straightforward to see that the conditional mean of b_i given d_i is likewise linear in d_i. Wu and Bailey (1988, 1989) noted that the model

$$E(b_i \mid d_i) \doteq \tilde{\psi}_0 + \tilde{\psi}_1 d_i$$

holds approximately for a broad class of random effects section models even when the normal assumption on d_i does not hold. More generally, Follman and Wu (1995) term these random effects models as shared parameters models, and have extended the methodology to the generalized linear model setting.

Mixture-Models. As we have seen, with random effects selection models the conditioning $f(d_i \mid Y_i)$ can be reversed to $f(Y_i \mid d_i)$; this concept underlies mixture models, which index the joint distribution of (Y_i, d_i) by $f(Y_i \mid d_i)\ f(d_i)$. This is in contrast to selection models which index the joint distribution by $f(Y_i)f(d_i \mid Y_i)$. Because the target of inference is ordinarily the parameters in $f(Y_i)$, the selection models seems more intuititive to many statisticians. In addition, the dependence of Y_i on d_i in $f(Y_i \mid d_i)$ is difficult to appreciate in a causal setting since it implies early values of Y_i may depend on subsequent dropout. Perhaps the best way to think about the mixture model is simply as an alternative method for factoring $f(Y_i, d_i)$ which offers some advantages over the selection model. If $f(Y_i \mid d_i) = f(Y_i^{\text{MIS}} \mid Y_i^{\text{OBS}})f(Y_i^{\text{OBS}} \mid d_i)$ then the missingness is non-informative. In the general case where $f(Y_i^{\text{MIS}}|Y_i^{\text{OBS}})$ depends on d_i as well, missingness is informative. When inference is desired for $f(Y_i)$ we must "mix" over the distribution of d_i, giving

$$f(Y_i) = \int f(Y_i \mid d_i)\ f(d_i)dd_i,$$

hence the term mixture models.

Random effects mixture models have been considered by Wu and Bailey (1988), Hogan and Laird (1997a) and Mori *et al.* (1992). These models offer some advantages over selection models, in that simpler estimation approaches are available, even with censored observations. For example, $f(d_i)$ may be estimated using the Kaplan-Meier estimator, and $f(Y_i^{\text{OBS}} \mid d_i, X_i)$ may be estimated from those with observed failure times, provided the censoring of d_i is unrelated to Y_i^{OBS}. Since here censoring means noninformative dropout **or** completers, this condition holds. With no censoring, estimation is very straightforward and can be done using standard techniques for longitudinal data analysis once one specifies a model for $f(Y_i \mid X_i, d_i, \beta, \theta)$. See Hogan and Laird (1997b) for details.

Little (1993) gives a general method for analyzing multivariate data with missing values which is appropriate when one can classify subjects on patterns of nonresponse, R_i. Here again, models for $f(Y_i \mid R_i = r_i)$ are posited, and constraints on $P(R_i|Y_i)$ are used in order to make the models identifiable. Little (1993) calls these pattern mixture models since potentially there is a different pattern for $f(Y_i)$ given each nonresponse pattern. A general review of mixture and selection models for dropouts in given in Little (1995).

9.3 Examples

In this section we will illustrate the use of selection models and mixture models to analyze non-ignorable data. Both examples are drawn from the literature (Fitzmaurice *et al.*, 1996, and Fitzmaurice and Laird, 2000); the reader is referred to these articles for more detail.

A Selection Model. This example involves bivariate repeated measures rather than longitudinal data, but the basic methodology is the same.

The data come from a sample survey of mental health among children. Many variables were assessed, including a measure of the child's emotional or internalizing behavior. This is measured by interview from an appropriate informant using a standardized scale, and then dichotomized for analysis. It is common in mental health surveys, especially of children, to use two or more informants. In this case, the identical information was gathered from both a parent and a teacher, hence the bivariate dichotomous outcome.

Because children could not participate without parental consent, there was little missing data in the parent assessment. However, there was considerable missing data in the teacher response. Many parents refused to give permission for the teacher assessment. In addition, many teachers simply failed to respond even when permission was given. There were 2,501 parents who returned valid questionnaires, but teacher assessments were not available for 43% of these. The investigators hypothesized that non-response was non-ignorable and that the cases of externalizing behavior were under represented among the observed teacher responses. This belief stemmed in part because some parents who refused permission for the teacher interview did so because they were concerned that the teacher would give the child "bad" ratings. This is of course not clear evidence of non-ignorable nonresponse, but does provide a rationale.

We first consider the analysis if we had complete data in order to specify a data model. One objective is simply to estimate the overall prevalence based on both the teacher and parent ratings. Secondly, we want to study the relationship of covariates to the behavior ratings, and the interaction between covariates and informant on ratings, adjusting for nonresponse. Here we will consider just two covariates, sex and a dichotomized measure of dissatisfaction with family life. These questions can be addressed by using a bivariate logistic regression model, with indicators for informant, the covariates and their interactions. The data for each subject can be thought of as arising from a 2×2 cross classification of each child on the two responses, and is thus multinomial. The basic model parameterizes the three independent parameters of the multinomial into two parameters corresponding to two the marginal log-odds and one corresponding to the odds ratio.

For example, consider the model which has main effects for informants and both covariates:

$$\text{logit} P(Y_{ij} = 1) = \beta_0 + \beta_1 I_j + \beta_2 X_{1i} + \beta_3 X_{2i}$$

where Y_{ij} denotes the response of the jth informant for the ith child, $I_1 = 1$ (for parent informant), and $I_2 = 0$ (for teacher informant), and

X_{1i} and X_{2i} denote the sex, and dissatisfaction variable for the ith child.

This model implies that there are informant differences in prevalence, as well as gender and dissatisfaction effects on prevalence, but that the informant effects do not depend on covariates. For simplicity, we assume that the OR is constant, i.e., not dependent on the covariates. This completes the specification of the data model. The model so defined, as well as ones including interaction terms, is fit using maximum likelihood.

To accommodate the missing teacher responses, we also use a logistic model for the response, non-response indicator, say, $R = 1$ if observed, and zero otherwise. The model can include as predictors the parent report, the potentially unobserved parent report, the covariates given in the data model as well as possibly other covariates of nonresponse, and their interactions. If the non-response model includes the teacher observation as a predictor, then the model is nonignorable, because non-response depends upon the potentially unobserved teacher response. If it includes only the parent observation and not the teacher, the model is MAR, and if it includes neither it is MCAR. Thus an advantage of this modeling approach is that different assumptions from the Rubin (1976) missing data paradigm can easily be fit and tested.

However, it is not possible to fit completely saturated models, because then the resulting estimation problem is ill conditioned, i.e., the parameters cannot be uniquely identified from the observed data. To see this, consider the case where there are no covariates in any model. Assume further that we fit a model for nonresponse that includes both parental and teacher observations, and their interaction and a data model that has an intercept and the informant effect. In this setting, the three parameters of the data model plus the four parameters of the nonresponse model completely determine the probabilities underlying the $2 \times 2 \times 2$ cross-classification of parent observation, teacher observation, and response indicator. However, we do not directly observe this $2 \times 2 \times 2$ table; rather, we observe the 2×2 table for parent and teacher observation given $R = 1$, and we observe the parent margin of the same 2×2 table for $R = 0$. Thus the saturated model is clearly not identifiable with the observed data.

One strategy which is possible with count data is to only fit models with degrees-of-freedom equal to directly observed cells; another is to evaluate the rank of the information matrix to determine estimability. Neither of these approaches work reliably but Glonek (1999) gives a straightforward approach to determining model identifiability for the categorical data setting.

Fitzmaurice *et al.* (1996) fit by ML a single data model with four different nonresponse models to the data. The data model assumes main effects for informant, gender and dissatisfaction, and the informant by dissatisfaction interaction; the nonresponse models all include the main effects of the covariates. The MCAR model includes no effects for parent or teacher observations. The MAR model includes parent observation. The third model replaces parent observation with teacher observation and the final model includes main effect for both parent and teacher observation. The MCAR and MAR models give identical results for the data model, as expected. Comparing the log-likelihoods suggests no evidence for preferring MAR to MCAR, and the estimated prevalence of internalizing disturbance is nearly the same for teachers and parents, and for the teacher complete cases (0.189).

The results for the MCAR and the two nonignorable models are given Table 1.

There is not much variation in the estimated coefficients of the data model across any of the nonresponse models; probably because there is little evidence from any model that nonresponse depends either on parent or teacher observation. Note that the large standard errors for the coefficient of teacher response in both of the nonignorable nonresponse models indicate that there is little information in the data about this coefficient. In addition, the standard errors for the data model are also very large when both teacher and parent observations are included in the model. The Fisher Information matrix for this model, evaluated at selected parameter values about the maximum was always non-singular, however.

In this example, the overall conclusions about prevalence and the effects of covariates on prevalence vary little under different assumptions on the nonresponse mechanism. It is important to keep in mind that the class of nonignorable models considered was limited, and using other models might show a bigger impact. Despite that, the results do not give credence to the hypothesis that the missingness was nonignorable. The models fit were kept simple to avoid identifiability problems, but even so, the estimation of the non-response model and also of the data model showed large levels of uncertainty.

Table 1

Measurement Parameters			Nonresponse Parameters		
Model/Parameter	Estimate	SE	Parameter	Estimate	SE
MAR					
Intercept	-1.472	.106	Intercept	.266	.071
Informant	-.280	.125	Y_P	.105	.106
Dissat	.417	.136	Dissat	.086	.083
Gender	-.411	.086	Gender	-.075	.081
Informant × Dissat	.549	.166			
Ln(OR)	.648	.156			
-2 Log-Likelihood = 7,097.84					
GOF: $X^2 = 7.31$, 10 df, ($p \approx .70$)					
NI					
Intercept	-1.927	.515	Intercept	.132	.160
Informant	.176	.521	Y_T	1.498	2.819
Dissat	.437	.133	Dissat	.042	.101
Gender	-.414	.086	Gender	-.045	.100
Informant × Dissat	.529	.163			
Ln(OR)	.646	.156			
-2 Log-Likelihood = 7,097.58					
GOF: $X^2 = 7.04$, 10 df, ($p \approx .72$)					
NI					
Intercept	-1.966	2.223	Intercept	.124	.539
Informant	.215	2.224	Y_T	1.733	14.963
Dissat	.438	.132	Y_P	-.010	.476
Gender	-.413	.086	Dissat	.039	.200
Informant × Dissat	.528	.163	Gender	-.044	.217
Ln(OR)	.643	.228			
-2 Log-Likelihood = 7,097.57					
GOF: $X^2 = 7.04$, 9 df, ($p \approx .63$)					

A Mixture Model. Here we use data from a clinical trial on contraception to illustrate the use of mixture models to handle nonignoragle nonresponse. In the trial, two dose levels of a drug were compared. The design called for injections at baseline and every three months for one year. The primary outcome of interest was the presence or absence of a side effect, amenorrhea, during the previous three months. The vector of outcomes for a person who completes the trial is a 4×1 vector of indicators of amenorrhea (Y_i). The trial objective was to compare trends in prevalence of amenorrhea in the two groups. However, there were many subjects who failed to return for subsequent scheduled injections, and it was thought that this might be related to the presence of amenorrhea (which is not observed) during the previous period. We describe an analysis presented in Fitzmaurice and Laird (2000).

With the mixture model, we need to specify a model for how the outcomes (Y_i) depend upon treatment group, and time of dropout, and another model for time of dropout. Since there are only three dropout times (excluding completing), and the dropout times are observed for everyone, it is easy to simply estimate the dropout probabilities for each time, stratified by treatment group, using the sample proportions. Denote these probabilities by π_{ik} where i indexes treatment group and k is time of dropout, $k = 1, \ldots, 4$, the last value indicating a completer, and $\pi_{i+} = 1$, $i = 1, 2$.

For the outcome model, we could again contemplate estimating $P(Y_{ij})$ as a function of time and treatment group, stratifying on dropout status. But it is immediately obvious that there are no data to estimate $P(Y_{i4})$ if dropout occurs on the third occasion, and similarly for Y_{i3}, Y_{i4} with dropout at the second occasion, etc. Another issue, which also arises with the selection model, is that if only a few dropouts occur, then even the identifiable parameters will be poorly estimated. A solution to both of these problems is to fit simpler model which indicates how outcome depends on dropout, and condition rather than stratify on dropout.

Fitzmaurice and Laird (2000) fit several models to investigate the sensitivity of results to modeling assumption. The first model assumes that $\text{logit}(P(Y_{ij} = 1))$ follows a quadratic in time, with different time coefficients for each dose group, and with intercept depending upon time of dropout (modeled as a dummy variable). That is, the effect of dropout is to shift up or down the overall level of amenorrhea, but not to change the shape. Thus effectively, the shape of the curve for early dropouts is extrapolated from the shape which is fit to the later dropouts and the completers. A more complex model allowed the shape to depend on both dropout status and treatment group, but was assumed to be the same for all dropouts (within a treatment group) at times 1, 2 and 3. They compared the results of these models with an ignorable model where $\text{logit}P(Y_{ij} = 1)$ is just quadratic in time, with different coefficients for each dose group. In principle, a complete likelihood could be specified for the Y_{ij}'s and ML estimation used for the regression parameters and any association parameters, but they used GEE to estimate the regression models and the empirical variance for the estimated standard errors.

The final step is to obtain estimates of the marginal $P(Y_{ij} = 1) = \mu_{ij}$, conditional on dose but not dropout. This is simply a summation,

$$\widehat{\mu}_{ij} = \sum_k \widehat{P}(Y_{ij} = 1 \mid \text{time, treatment, dropout} = k)\widehat{\pi}_{ik},$$

where the estimated probabilities of $P(Y_{ij} = 1)$ are the anti-logits from the fitted regression models.

The results of model fitting are given below for the difference in mean vectors, say $\delta_j = \mu_{hj} - \mu_{lj}$ for the high and low groups at time point j. There is very little variation in the estimated δ_j, for any of the three models, except time 4 where the most complicated model estimates half the difference of the other two, but no model suggests that the treatment groups are different at the end of the trial.

Generalized Linear Mixture Models

Marginal rates of amenorrhea at the four study time pointime under three different modelling assumptions about dropout

Model	Time j	$\widehat{\delta}$ Difference	S.E.	Z	p
MCAR	1	0.017	0.023	0.73	0.4629
	2	0.089	0.025	3.54	0.0004
	3	0.109	0.030	3.65	0.0003
	4	0.052	0.036	1.46	0.1437
Dropout affects level	1	0.016	0.023	0.72	0.4734
	2	0.091	0.026	3.57	0.0004
	3	0.112	0.030	3.71	0.0002
	4	0.053	0.035	1.52	0.1275
Dropout affects slope	1	0.016	0.023	0.70	0.4864
	2	0.100	0.029	3.39	0.0007
	3	0.105	0.041	2.55	0.0108
	4	0.025	0.056	0.44	0.6614

Summary. With both the selection model and the mixture model examples, the use of several different models which reflect different assumptions about dropout are used to explore the sensitivity of results to model assumptions. We emphasize that this is perhaps the best use of the various methods available for handling nonignorable dropouts. In both of our examples, investigators felt that the nonresponse process was nonignorable, but in neither case did our models give evidence of that. Bear in mind that there are many possible models that one can fit, however, and some may well give different results. An alternative to estimating the model parameters would be to fix them at prespecified values. This has the advantage of ameliorating the problems of small number of dropouts and lack of identifiably.