

# Ignorable dropout in longitudinal studies

BY MARIA ELEANOR V. TIPA, SUSAN A. MURPHY  
*Department of Statistics, The Pennsylvania State University,  
University Park, Pennsylvania 16802, U.S.A.*

AND DIANE K. MCLAUGHLIN  
*The Population Research Institute, The Pennsylvania State University,  
University Park, Pennsylvania 16802, U.S.A.*

## SUMMARY

This paper provides a concise definition for ignorable dropout. This is done primarily from a frequentist perspective. The definition of ignorable dropout depends on both the population of inference and the type of statistical methodology used for inference. Different types of dropouts are described and compared to those found in the literature. Ignorability conditions are then given for the following types of inference: Likelihood-based Inference, Preservation of Marginal Moments and Preservation of Conditional Moments.

*Some key words:* Censoring, Missing Data.

## 1. INTRODUCTION

In a longitudinal study, measurements are taken on a unit or subject repeatedly at different time points. If measurements are taken  $n$  times, then the complete vector of measurements on the  $i^{\text{th}}$  unit is  $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in})$ . It is not uncommon for longitudinal data to have missing values, that is, for some  $i$ , the whole  $\mathbf{Y}_i$  vector will not be completely observed. A special case of this is when subjects leave the study prematurely, that is, if  $Y_{ij}$  is missing then all the succeeding  $Y_{ik}$ 's,  $k > j$ , are also missing. This special case is commonly called a *dropout* ((Diggle, 1989), (Diggle & Kenward, 1994), (Heyting et al, 1992), (Little,1995)).

Dropouts cause the data to be unbalanced and there are statistical procedures that can handle unbalanced data. However caution is required in using these techniques in the presence of dropouts. This is because the dropout may cause bias. For instance, in

a clinical trial patients who recover may tend to drop out more than those who do not. Then the group with more cured subjects will also have more dropouts. Analysis of the observed data will then be biased against the group with more recovered patients.

However there are cases where valid inference still results even in the presence of dropout. Rubin (1976) calls such dropouts *ignorable*, otherwise they are *nonignorable*. Since different types of statistical methodology require different conditions for inference to be valid, the definition of *ignorability* depends on the statistical procedure used. Ignorability has been examined by many authors ((Rubin, 1976), (Laird, 1988), (Little & Rubin, 1987), (Diggle & Kenward, 1994), (Little, 1995)). Except for the work by Rubin (1976), the complex relationship between dropout, the population of inference and the statistical methodology is not emphasized.

In this paper we discuss this complex relationship and give sufficient conditions for for ignorable dropout for three statistical methodologies. This is done from a frequentist perspective. In Section 2 the definition of dropout is related to the population of inference. A unit that leaves the study prematurely is not necessarily a dropout – it should be evaluated as being a dropout or not relative to the population of inference. In section 3 different types of dropout are defined. Section 4 provides ignorability conditions for the following types of frequentist inference: 1) Likelihood-Based Inference, 2) Preservation of Marginal Moments and 3) Preservation of Conditional Moments.

## 2. DROPOUT AND THE POPULATION OF INFERENCE

We say that a subject leaving a study prematurely is a dropout if and only if following the departure the subject remains in the population of inference. Note the explicit reference to the population. We illustrate this definition of dropout using three studies: the Panel Study of Income Dynamics (PSID) (Hill, 1992), the Multicenter AIDS Cohort Study (MACS) (Kaslow et al., 1987) and a clinical trials example (Heyting et al., 1992).

Suppose that we are interested in explaining transitions into and out of poverty

for the nonimmigrant, noninstitutionalized elderly, those aged 55 or over. That is, our population of inference is the set of all nonimmigrants, aged 55 or over, residing in the United States and not living in institutions. The PSID is an ideal data set for this purpose since the subset of the elderly in the PSID sample, when properly weighted, is representative of the above population. Further, sample families have been interviewed annually since 1968 to collect information on variables such as sources of income, employment information, work hours, geographic mobility and other demographic variables (User Guide to the PSID, 1984). This makes it possible for us to trace the economic history of the respondents since entry into the study.

The PSID loses some respondents for a variety of reasons, including: 1) change of address - unknown new address; 2) refusal to be interviewed; 3) institutionalization which can include going into service, moving into a nursing home, moving into a dormitory, etc., and; 4) death (PSID: Tapes and Procedures, 1986). Whether or not these losses should be considered dropouts depends on the the population given above. Certainly individuals who were not interviewed for reasons like residence change or refusal to be interviewed are dropouts. We know that these respondents can be in one of two states: in poverty or not in poverty but the state is unknown.

A more complex question is whether people lost from the study due to death are dropouts. In our population of inference, death is a natural occurrence. Therefore, those who die are not dropouts. Similarly, a person who goes into an institution ceases to be a member of our population of inference and thus, we do not lose any information once he enters an institution. In both cases a person leaving the study due to death or institutionalization to a nursing home is not a dropout since we want to make generalizations about the poverty experiences of the noninstitutionalized elderly in the U.S. Death and institutionalization are conditions or states a person can be in aside from being in poverty or not being in poverty. In survival analysis, these are called competing risks where for a given period of time, a person can be in one of different possible states. We see then that death or institutionalization signifies

the end of a complete observation and that there is no loss of information.

Next consider a conceptual population of people, aged 55 and older, who reside in the U.S. and for whom institutionalization is not permitted/possible. Then if in our sample an individual becomes institutionalized, this is a dropout. This is because we do not have information on the person's poverty status which would have been available had they not been permitted or allowed to enter an institution.

On the other hand suppose our population of inference is the set of all people, aged 55 and older who reside in the U.S. Is poverty a relevant measure for people in institutions – that is, can a person be poor and in an institution at the same time? If an institutionalized person can be poor or not poor, then after a person in our sample becomes institutionalized, we no longer have complete information about his poverty status. He is a dropout. If however a person cannot be poor or not poor in an institution, institutionalization is then a competing risk. In this case, we do not lose any information if someone leaves the study due to institutionalization. Therefore, he is not a dropout.

Let us now consider the MACS example. The MACS is a comprehensive longitudinal study of human immunodeficiency virus (HIV) infection among homosexual men who may or may not be HIV-positive at the time of entry into the study. It has been conducted in four metropolitan areas (Boston, Pittsburgh, Los Angeles and Baltimore/Washington, DC) starting in 1984 (Kaslow et al., 1987). Demographic variables, medical history, drug use and sexual practices were collected at entry into the study along with hematological variables. All subjects are scheduled for reevaluation twice a year while some subjects are recalled every three months. Hoover et al., (1993) cited the onset of AIDS as a primary reason for subjects leaving the study prematurely.

Suppose that we are interested in the population of homosexual men who are HIV-positive. We wish to investigate the progression of the HIV infection as measured by CD4 cell count. A subject who leaves the study will have missing CD4 counts and is

thus a dropout. This includes the case of a subject leaving the study due to the onset of AIDS since we are interested in CD4 cell counts even after the onset of AIDS. On the other hand, a person who dies during the study is not a dropout. In our population of inference, death is a natural occurrence. Similar to the death example in the PSID given previously, death signifies the end of a complete observation.

Consider now the case when we wish to make inferences on the conceptual population of HIV-positive homosexual men where the only cause of death is AIDS (this is a plausible population if the probability of dying from other causes is negligible). A person who leaves the study is a dropout. Likewise, a person who dies of causes not related to AIDS is a dropout. Death from AIDS, however, is a natural occurrence with respect to our population of inference and is thus not a dropout.

This is an example in which some authors have modelled both the outcome variable (CD4 count) that is measured periodically and the survival time for an event (onset of AIDS). Hogan and Laird (1996) prefer to use different terms to describe processes that lead to incomplete observations on the survival time and the outcome variable – they use *dropout* for the outcome variable and *censoring* for the survival time.

Suppose however that we are interested in the conceptual population of people who have HIV infection where there is no attrition due to death from AIDS. Then deaths due to AIDS are dropouts since information on their CD4 counts would have been known had they not died. In this setting, De Gruttola and Tu (1994) modelled both the disease progression and the survival time to account for the missing CD4 counts due to deaths from AIDS.

Let us now look at clinical trials. Suppose that two treatments are being compared, treatment A and treatment B, and subjects are evaluated periodically. Subjects may leave the study or die before the study's end or they may switch to a different treatment (something other than A or B). Subjects who leave the study are not followed up but those who switch to another treatment are still monitored. Schwartz and Lellouch (1967) distinguished between an *explanatory* analysis and a *pragmatic*

analysis. In an explanatory analysis, the primary goal is to compare the effect of the two treatments under controlled conditions. A pragmatic analysis, on the other hand, aims to be able to make inferences on the best possible treatment or combinations of treatments under real-life conditions.

Say that we are interested in the population of patients receiving the particular treatment (A or B) under investigation for the entire study length. We wish to determine their response over time to this treatment – that is, we want to do an explanatory type of analysis. A subject who leaves the study is of course a dropout. A subject who switches to a different treatment is similar to someone leaving the study in the sense that in both cases, we do not have the information of what their response would have been had they continued on the treatment. A person who dies during the study is not a dropout since this is a natural occurrence in our population of inference. If we are interested however in the conceptual population of these patients where the only possible cause of death is the disease being treated, then deaths due to causes other than this disease are dropouts. A death due to this disease signifies the end of a complete observation and thus is not a dropout.

Consider now the case where we are interested in the population of patients receiving the particular treatment under investigation (A or B) as their initial treatment. We wish to investigate their response to the initial treatment and to subsequent treatments thereafter – that is, we want to do a pragmatic analysis. This is also called an “intention-to-treat” analysis ((Fisher et al., 1990), (Hogan & Laird, 1995)). A subject who leaves the study will result in missing information and is thus a dropout. A subject who switches to a different treatment is not a dropout with respect to this population. Similar to the explanatory type of analysis, death is not a dropout since this is a natural occurrence in our population of inference. However in the conceptual population of patients where death can only be due to the disease being treated, deaths due to other causes are dropouts. Death due to the disease is not a dropout.

As we have seen a sample may be used for inference about more than one pop-

ulation. And a dropout for one population of inference may not be a dropout for a different population of inference. It is important to first determine the population of inference and then define dropout for this population.

### 3. TYPES OF DROPOUT

Suppose that the units in our study are observed in discrete time,  $(t_1, \dots, t_K)$ . The ideal data for a subject is represented by  $\mathbf{Y} = (Y_1, \dots, Y_K)$  and  $\mathbf{X} = (X_1, \dots, X_K)$  corresponding to the  $K$  time points. We assume that for any  $k$  the distribution of  $(Y_1, \dots, Y_k)$  given  $(X_1, \dots, X_k)$  is independent of  $(X_{k+1}, \dots, X_K)$ . In other words,  $X$  may be a function of time and baseline variables, which are completely observed at time  $t_1$ , and/or  $X$  is an external covariate. Conditioning on past outcomes, the ideal data likelihood function,  $L(\theta)$  for one subject can be expressed as

$$L(\theta) = f_{Y_1}(Y_1|\mathbf{X};\theta) \prod_{k=2}^K f_{Y_k}(Y_k|Y_1, \dots, Y_{k-1}, \mathbf{X};\theta),$$

where  $f_{Y_1}(y_1|\mathbf{X};\theta)$  is the marginal density of  $Y_1$  given  $\mathbf{X}$  and the conditional density of  $Y_k$  given the past outcomes  $Y_1, \dots, Y_{k-1}$  and  $\mathbf{X}$  is  $f_{Y_k}(y_k|Y_1, \dots, Y_{k-1}, \mathbf{X};\theta)$ .

We now consider the case where dropout is possible. In order to focus on a typical subject's response and dropout, we assume that observations on the subjects are independent and identically distributed. We also assume that  $\mathbf{X}$  is always observed. For one subject, the complete data can be represented by  $\{(Y_1, \mathbf{X}), (Y_2, D_2), \dots, (Y_K, D_K)\}$  where  $D_k = I(D > k)$  and  $D$  is the time of dropout. The complete data likelihood for one subject is given by

$$f_{Y_1}(Y_1|\mathbf{X};\theta) \prod_{k=2}^K f_{Y_k, D_k}(Y_k, D_k|(Y_j, D_j)_{j < k}, \mathbf{X}; \psi)$$

where  $\psi = (\theta, \phi)$  and  $\phi$  contains parameters that describe the conditional distribution of  $(D|Y, \mathbf{X})$ . Note that in this representation, there may be an overlap between  $\theta$  and  $\phi$ .

In reality, when  $D_k = 0$ ,  $Y_k$  is not observed. Let  $\tilde{Y}_k = Y_k$  if  $D_k = 1$  and  $\tilde{Y}_k = *$  otherwise. The incomplete data can then be represented as  $\{(Y_1, \mathbf{X}), (\tilde{Y}_2, D_2), \dots, (\tilde{Y}_K, D_K), \}$ . The incomplete data likelihood function is

$$L(\psi) = f_{Y_1}(\tilde{Y}_1 | \mathbf{X}; \theta) \prod_{k=2}^K \{f_{Y_k, D_k}(\tilde{Y}_k, 1 | \text{past}_k; \psi)\}^{D_k} \prod_{k=2}^K \{f_{D_k}(0 | \text{past}_k; \psi)\}^{1-D_k}$$

where  $\text{past}_k = \{\mathbf{X}, (Y_j)_{j < k}, D \geq k\}$ . This is because when  $D_k = 1$ , the joint distribution of  $(\tilde{Y}_k, D_k)$  is the same as that of  $(Y_k, D_k)$  and when  $D_k = 0$ ,  $Y_k$  is not observed so we merely use the distribution of  $D_k$ .

We now describe three conditions on dropout that appear frequently in the literature. In the following section these conditions on dropout will be used to establish ignorability for different methods of inference. The conditions will be interpreted in the context of the population of nonimmigrant, noninstitutionalized people aged 55 or over (people in late mid-life through old-age) who were residing in the U.S. in 1968. As we shall see, conditional independence is a statement about subgroups of this population. For the interpretations that will be given later on, we will assume, for simplicity, that the only reason for dropout is emigration. The response is whether a person is in poverty or not.

Moreover, unless otherwise indicated, the conditions that will be given are defined for all possible samples from the population.

One condition, which we will call *independent dropout I*, is that for a person still in the study up to time  $t_k$  the probability density of his potential response  $Y_k$  is the same as the assumed distribution in the case when dropout is not a possibility. Kalbfleisch and Prentice (1980, p. 120) and Andersen et al. (1988, p. 30) referred to an analogous condition, *independent censoring*, in the context of survival analysis for continuous lifetimes. This condition can be quantified as

$$Y_k \text{ is independent of } \{D \geq k\} \text{ given } (\mathbf{X}, Y_j, j < k)$$

for  $k = 1, \dots, K$ . Or in terms of the densities,

$$f_{Y_k}(y_k | \text{past}_k; \psi) = f_{Y_k}(y_k | \mathbf{X}, Y_j, j < k; \theta) \tag{1}$$

for  $k = 1, \dots, K$ . Note also that (1) implies that the conditional density of  $(Y_k | \mathbf{X}, Y_j, j < k, D \geq k)$  depends only on  $\theta$ .

We now illustrate this condition using the population described earlier. Consider the group  $S_1$  of people with a similar past, say college graduate males who have not been poor from age 55 up to age  $t_k$ , and let  $P_k$  be the group of males in  $S_1$  who are poor at age  $t_k$ . Suppose that the proportion of males in poverty at age  $t_k$  in  $S_1$  is 0.10. That is,  $\frac{n(P_k)}{n(S_1)} = 0.10$ . Now, divide  $S_1$  into  $S_2$  and  $S_1 \setminus S_2$ , where  $S_2$  contains those who have not emigrated before  $t_k$  and  $S_1 \setminus S_2$  contains those who have emigrated before  $t_k$ . If the independent dropout I assumption is to be satisfied, then the proportion of poor men in  $S_2$  should also be 0.10, that is,  $\frac{n(P_k \cap S_2)}{n(S_2)} = 0.10$ . We illustrate this in Figure 1. Note that this also implies that the proportion of poor men at  $t_k$  in  $(S_1 \setminus S_2)$  is also 0.10. In effect, we are assuming here that those who have not emigrated before  $t_k$  are as likely to be poor at age  $t_k$  as those who have emigrated before  $t_k$ .

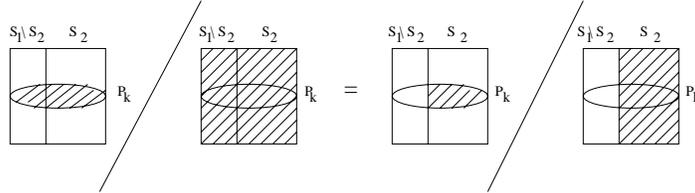


Figure 1: Independent Dropout I

Suppose that men unwilling to emigrate are more likely to become poor. In this case, relative to group  $S_1$ , group  $S_2$  would be overly composed of men resistant to emigration and hence more likely to be poor. The independent dropout I assumption would not hold in this case.

A second condition is the *random dropout* condition discussed by Diggle and Kenward (1994, p. 53). They defined this as that the probability of a dropout at time  $t_k$  depends only on the previous observations and not on the observation at  $t_k$ . Andersen et al. (1988, p. 30) described an analogous condition that the censoring process should depend (in a functional sense) only on the past and not on future events. They referred to censoring processes that satisfy this condition as *predictable*

censoring processes.

Mathematically, we can express the random dropout condition as

$Y_k$  and  $D_k$  are conditionally independent given  $\text{past}_k$

for  $k = 1, \dots, K$ . In terms of densities:

1.

$$f_{D_k|Y_k}(1|\text{past}_k, Y_k; \psi) = f_{D_k}(1|\text{past}_k; \psi) \quad (2)$$

or equivalently,

$$f_{Y_k|D_k}(y_k|\text{past}_k, D_k = 1; \psi) = f_{Y_k}(y_k|\text{past}_k; \psi), \quad (3)$$

for  $k = 1, \dots, K$ .

2. Given the past, the distribution of the potential response at time  $t_k$  is the same for someone who drops out at time  $t_k$  and someone who does not drop out at time  $t_k$ . In other words, we are saying that after accounting for their pasts, those who drop out are not that different from those who do not. This can be expressed as

$$f_{Y_k|D_k}(y_k|\text{past}_k, D_k = 0; \psi) = f_{Y_k|D_k}(y_k|\text{past}_k, D_k = 1; \psi) \quad (4)$$

for  $k = 1, \dots, K$ .

It can be shown that (2), (3) and (4) are equivalent to each other.

We now interpret Equation (4) using the elderly population. Define a subgroup in  $S_2$ , group  $E_k$ , composed of men who emigrate at  $t_k$ . Equation (4) requires that the proportion of poor men in  $S_2$  among those who do not emigrate at  $t_k$  is the same as the the proportion of poor men in  $S_2$  among those who emigrate at  $t_k$ . That is,  $\frac{n(P_k \cap E_k' \cap S_2)}{n(E_k' \cap S_2)} = \frac{n(P_k \cap E_k \cap S_2)}{n(E_k \cap S_2)}$ . Thus, among the men who have not emigrated prior to  $t_k$  those who emigrate at  $t_k$  are as likely to be poor at  $t_k$  as those who do not emigrate at

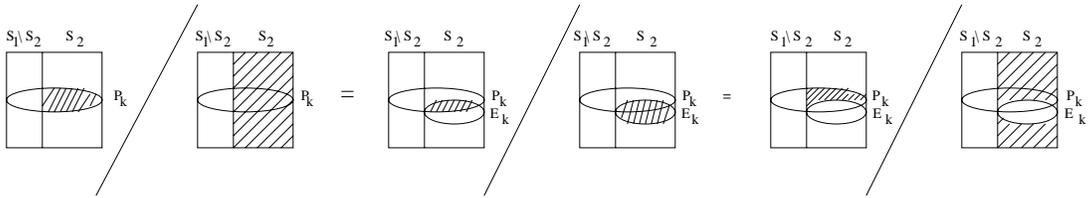


Figure 2: Random Dropout

$t_k$ . We show this in the second and third sets of diagrams in Figure 2. By Equation (3), we have the equality with the first diagram.

Diggle and Kenward (1994) called the dropout *informative* if the random dropout condition is not satisfied, that is, if the probability of dropout at time  $t_k$  depends on the potential response  $Y_k$ . In estimating the rate of change over time in a continuous variable in a random effects model, Wu and Carroll (1988), Wu and Bailey (1989) and Schluchter (1992) call the censoring process *informative* if the censoring probability for each individual is related to the random effects for that individual. Little (1995) distinguishes between these two uses of the word *informative* by calling the first *nonignorable outcome-based dropout* in which the dropout probability depends on the missing values of the response variable and the second *nonignorable random-coefficient-based dropout* where the dropout probability depends on the random effects. From a frequentist perspective, nonignorable random-coefficient-based dropout is one way in which nonignorable outcome-based dropout can occur. That is, dropout probability is related to the potential response via the random effect.

Another condition we will use to establish ignorability of the dropout is the *independent dropout II* condition,

$$Y_k \text{ is independent of } \{D > k\} \text{ given } (\mathbf{X}, Y_j, j < k)$$

for  $k = 1, \dots, K$ . In terms of densities,

$$f_{Y_k|D_k}(y_k|\text{past}_k, D_k = 1; \psi) = f_{Y_k}(y_k|\mathbf{X}, Y_j, j < k; \theta) \quad (5)$$

for  $k = 1, \dots, K$ . This is one possible quantification of Diggle and Kenward's (1994) assumption that "if an experimental unit is still in the study at time  $t_k$  its associated

sequence of measurements  $\tilde{Y}_j : j = 1, \dots, k$  follows the same joint distribution as that of  $Y_j : j = 1, \dots, k$ . Diggle and Kenward do not quantify this statement.

We are assuming here that given the past and that the subject has not dropped out at the present time, the distribution of his response at the present time is the same as in the case when there is no possibility of a dropout. In the population described earlier, we again consider the group,  $S_1$ , of college graduate males who have not been poor from age 55 up to age  $t_k$  and the following subgroups in  $S_1$ :  $S_2$ , composed of men who have not emigrated before  $t_k$ ; and  $P_k$ , composed of men who are poor at  $t_k$ .

Suppose that the proportion of poor men at age  $t_k$  in  $S_1$  is equal to 0.10. That is,  $\frac{n(P_k)}{n(S_1)} = 0.10$ . Now, we consider only those men who do not emigrate up to and including at age  $t_k$  ( $E'_k \cap S_2$ ). Under Condition (5), the proportion of poor men in this set is also 0.10. That is,  $\frac{n(P_k \cap E'_k \cap S_2)}{n(E'_k \cap S_2)} = 0.10$ . This is shown in Figure 3.

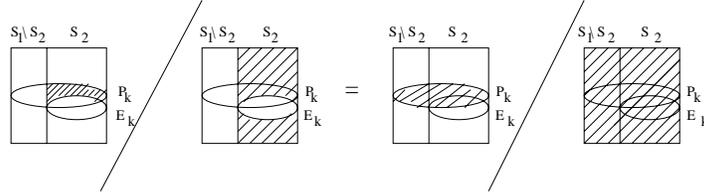


Figure 3: Independent Dropout II

From Figures 1, 2, 3 we see that any two conditions imply the third. If we interpret Diggle and Kenward's qualitative assumption as independent dropout II (5) then, together with their random dropout condition (2), they are implicitly assuming independent dropout I (1) also. It should be noted however that no one condition implies either of the two other conditions. For instance, suppose that  $Y_1, Y_2, Y_3$  are independent uniform random variables defined over  $(0, 1)$  and that  $f_{D_2|\mathbf{Y}}(1|\mathbf{Y}) = I(y_2 < 0.5)$  and  $f_{D_3|\mathbf{Y}}(1|\mathbf{Y}, D_2 = 1) = I(y_3 < 0.5, y_2 < 0.5)$ . In this example, independent dropout I holds but neither random dropout nor independent dropout II hold.

These conditions are similar to the conditions used in the literature for general missing data patterns. Rubin (1976) defined missing data as *missing at random*

(MAR) if the probability of the observed pattern of missing data does not depend on the missing  $y$  values. If we apply this to longitudinal data where a dropout occurs at time  $t_d$  and we observe  $\mathbf{X} = \mathbf{x}, (\tilde{Y}_1 = \tilde{y}_1, \dots, \tilde{Y}_{d-1} = \tilde{y}_{d-1})$ , we can interpret this condition as

$$P(D = d | \mathbf{X}, Y_1, \dots, Y_K; \phi) |_{\mathbf{x}=\mathbf{x}, Y_1=\tilde{y}_1, \dots, Y_{d-1}=\tilde{y}_{d-1}} = P(D = d | \mathbf{X}, Y_1, \dots, Y_{d-1}; \phi) |_{\mathbf{x}=\mathbf{x}, Y_1=\tilde{y}_1, \dots, Y_{d-1}=\tilde{y}_{d-1}} \quad (6)$$

for the observed time of dropout,  $d$ , the observed past,  $(\mathbf{x}, \tilde{y}_1, \dots, \tilde{y}_{d-1})$  and for all unobserved  $Y_d, \dots, Y_K$ . We will call this condition as *Bayesian MAR*. Some authors ((Laird, 1988), (Heyting et al., 1992)) use (6) and assume it to hold for all  $d = 2, \dots, K$ , all  $(\tilde{y}_1, \dots, \tilde{y}_{d-1})$  and all  $(Y_d, \dots, Y_K)$ . We will call this condition as *frequentist MAR*. That is, Bayesian MAR is (6) stated only for the actual sample that has been observed and all unobserved  $Y_d, \dots, Y_K$  values while frequentist MAR is (6) stated for all possible samples that may be collected from the population. Laird (1988) describe this condition as that “missingness or nonresponse mechanism should depend only on the observed values and not on the unobserved ones” while Heyting et al. (1992) describe it as “conditional on the observed responses, missingness occurs at random”.

As is shown by Robins et al. (1995), frequentist MAR is equivalent to the following strengthening of random dropout:

$$D_k \text{ is independent of } (Y_k, \dots, Y_K) \text{ given } \text{past}_k$$

for  $k = 2, \dots, K$  or equivalently,

$$f_{D_k}(1 | \text{past}_k; \psi) = f_{D_k | Y_k, \dots, Y_K}(1 | \text{past}_k, Y_k, \dots, Y_K; \psi), \text{ for } k = 2, \dots, K. \quad (7)$$

However, Bayesian MAR is not equivalent to (7).

Heyting et al. (1992) assume Equation (7) in estimating the mean response at assessment time  $t_k$  in clinical trials, where the mean is weighted by the probability

of staying in the trial,  $f_{D_k}(1|\text{past}_k)$ . Robins et al. (1995) also proposed a class of weighted estimating equations under the condition that

$$f_{D_k}(1|D_{k-1} = 1, \bar{W}_k) = f_{D_k}(1|D_{k-1} = 1, \bar{W}_k, \mathbf{Y}), \quad \text{for } k = 2, \dots, K, \quad (8)$$

where  $\bar{W}_k = \{\mathbf{X}, V_0, Y_0, \dots, V_{k-1}, Y_{k-1}\}$  where  $V$  is a matrix of time-dependent covariates. They also assume that  $f_{D_k}(1|D_{k-1} = 1, \bar{W}_k)$  is known up to a parametric form. In our simple formulation, there is no  $V$ . In this case, (8) is equation (7).

One may show that frequentist MAR (6 or 7) implies that independent dropout I (1) holds. However the independent dropout I condition does not imply frequentist MAR. One way to interpret the independent dropout I condition is that the conditional probability of dropping out at any time *prior* to time  $t_k$  is independent of  $Y_k$  given  $(Y_1, \dots, Y_{k-1})$ . We see that this can be satisfied even if the conditional probability of dropping out *at* time  $t_k$  given the past depends on  $Y_k$  which violates frequentist MAR.

Frequentist MAR (6 or 7) also implies that random dropout (2) holds. However, if

$$D_k \perp Y_{k+1}, \dots, Y_K \text{ given } \text{past}_k, Y_k \quad \text{for } k = 2, \dots, K. \quad (9)$$

then (7) is equivalent to (2). It would seem that (9) would be satisfied in many practical cases. Thus, in situations in which (9) is plausible, frequentist MAR is not substantially stronger than random dropout. Since independent dropout I and random dropout imply independent dropout II, it follows that frequentist MAR also implies the independent dropout II condition (5).

#### 4. IGNORABILITY OF THE DROPOUT

We call the dropout *ignorable* if and only if the mechanism of inference for the ideal data (but possibly unbalanced) is valid for the incomplete data subject to dropout. In likelihood inference, inference for parameters in marginal means and inference for parameters in conditional means, large sample properties, such as consistency

and asymptotic normality of the estimator of the parameter are used for hypothesis testing and construction of confidence intervals. In particular, it is necessary that the estimating function for  $\theta$  be unbiased in order for the estimator of  $\theta$  to be consistent and asymptotically normal. We focus on this minimal property in defining ignorable dropout.

Valid inference on the parameter  $\theta$  using likelihood methods depends on the correct specification of the likelihood function up to proportionality constants not depending on  $\theta$ . Then, in general, the score function evaluated at the true value  $\theta$  will have mean zero (will be an unbiased estimating function).

Suppose we observe,  $\mathbf{X}, D$ , and  $\tilde{Y}_1, \dots, \tilde{Y}_{D-1}$ . Then if we pretend that  $D$  is a constant and that there are only  $D - 1$  observations, we will use the function

$$L^C(\theta) = f_{Y_1}(\tilde{Y}_1|\mathbf{X}; \theta) \prod_{k=2}^{D-1} f_{Y_k|Y_1, \dots, Y_{k-1}}(y_k|\mathbf{X}, y_1, \dots, y_{k-1}; \theta)|_{y_1=\tilde{Y}_1, \dots, y_k=\tilde{Y}_k} \quad (10)$$

and  $\frac{\partial}{\partial \theta} \log L^C(\theta) =$

$$\frac{\partial}{\partial \theta} \log f_{Y_1}(\tilde{Y}_1|\mathbf{X}; \theta) + \sum_{k=1}^K D_k \frac{\partial}{\partial \theta} \log f_{Y_k|Y_1, \dots, Y_{k-1}}(y_k|\mathbf{X}, y_1, \dots, y_{k-1}; \theta)|_{y_1=\tilde{Y}_1, \dots, y_k=\tilde{Y}_k}$$

as one subject's contribution to the likelihood and score function respectively. In this case we are *ignoring the dropout*. Note that using this ignores the fact that the number of observed values of  $Y$  depends on the random variable  $D$ . The correct likelihood to use is the incomplete data likelihood,

$$L(\psi) = f_{Y_1}(\tilde{Y}_1|\mathbf{X}; \theta) \prod_{k=2}^K f_{Y_k, D_k}(\tilde{Y}_k, 1|\text{past}_k; \psi)^{D_k} \prod_{k=2}^K f_{D_k}(0|\text{past}_k; \psi)^{1-D_k}, \quad (11)$$

where  $\text{past}_k = (\mathbf{X}, Y_1, \dots, Y_{k-1}, D \geq k)$ ,  $\psi = (\theta, \phi)$  and  $\phi$  contains parameters that describe the conditional distribution  $(D|Y, \mathbf{X})$ .

If we can factor  $L(\psi)$  into  $L^C(\theta)$  and some other factor not involving  $\theta$ , then the correct score function for  $\theta$  will be  $\frac{\partial}{\partial \theta} L^C(\theta)$  summed over all subjects. In this case, we say that the dropout is *ignorable*. We now give two settings for which the dropout is ignorable.

**Theorem 1** *The dropout is ignorable under likelihood-based inference if,*

1. *the independent dropout II condition holds for  $k = 1, \dots, K$ , and;*
- 2.

$$f_{D_k}(1|\text{past}_k; \psi) \text{ is functionally independent of } \theta \quad (12)$$

*holds for  $k = 1, \dots, K$ .*

Equation (12) requires that the conditional probability of a dropout given the past should not contain any information about the parameter of interest.

The above theorem is proved by factoring  $L(\psi)$  as

$$f_{Y_1}(\tilde{Y}_1|\mathbf{X}; \theta) \prod_{k=2}^K f_{Y_k|D_k}(\tilde{Y}_k|\text{past}_k, D_k = 1; \psi)^{D_k} f_{D_k}(1|\text{past}_k; \phi)^{D_k} f_{D_k}(0|\text{past}_k; \phi)^{1-D_k}.$$

Under independent dropout II,  $f_{Y_1}(\tilde{Y}_1|\mathbf{X}; \theta) \prod_{k=2}^K f_{Y_k|D_k}(\tilde{Y}_k|\text{past}_k, D_k = 1; \psi)^{D_k}$  is equal to  $L^C(\theta)$ . In this formulation,  $L^C(\theta)$  is a partial likelihood. Cox (1975) showed that under the usual regularity conditions, the score function from the partial likelihood has similar asymptotic properties as the score function from the full likelihood. This implies then that even if (12) is not satisfied, that is, even if the conditional probability of a dropout given the past depends on  $\theta$ , we can still make valid large sample frequentist likelihood inference on  $\theta$  based on  $L^C(\theta)$  alone. However there will be a loss in efficiency if we use the partial likelihood instead of the full likelihood.

Note that for a particular observed sample the likelihood will be correctly specified, if Equation (5) and Equation (12) hold only for the particular observed data and all unobserved  $Y$  values. Thus, for direct-likelihood inference which Rubin (1976) defines as inference that “results solely from ratios of the likelihood function”, assuming (5) and (12) for the observed data and all unobserved data is sufficient for ignorability. Rubin’s conditions for ignorability under direct-likelihood inference are that the missing data are Bayesian MAR and that  $\theta$  should be distinct from  $\phi$ . However to consider the properties of estimators derived from the likelihood function like consistency and asymptotic distribution theory, we assume these conditions for all

possible samples from our population. Laird (1988) and Heyting et al. (1992) use the frequentist MAR condition (6/7) for ignorability and implicitly assume that  $\theta$  and  $\phi$  are distinct. We have shown in the previous section that frequentist MAR (6/7) implies both independent dropout I and independent dropout II.

An alternate condition for ignorability is:

**Theorem 2** *The dropout is ignorable under likelihood-based inference if,*

1. *the independent dropout I condition holds for  $k = 1, \dots, K$ , and;*
- 2.

$$f_{D_k|Y_k}(1|\text{past}_k, Y_k; \psi) \text{ and } f_{D_k}(1|\text{past}_k; \psi) \text{ are functionally independent of } \theta \quad (13)$$

*holds for  $k = 1, \dots, K$ .*

This theorem results if we express  $L(\psi)$  as

$$L(\psi) = f_{Y_1}(\tilde{Y}_1|\mathbf{X}; \theta) \prod_{k=2}^K f_{D_k|Y_k}(1|\text{past}_k, Y_k = \tilde{Y}_k; \psi)^{D_k} f_{Y_k}(\tilde{Y}_k|\text{past}_k; \psi)^{D_k} f_{D_k}(0|\text{past}_k; \psi)^{1-D_k}.$$

Under independent dropout I,  $f_{Y_1}(\tilde{Y}_1; \theta) \prod_{k=2}^K f_{Y_k}(\tilde{Y}_k|\text{past}_k; \psi)^{D_k}$  is equal to  $L^C(\theta)$ . In this formulation then,  $L^C(\theta)$  is not a partial likelihood. This means then that if  $\frac{L(\psi)}{L(\theta)}$  depends on  $\theta$ , we can not make likelihood inferences on  $\theta$  based on  $L^C(\theta)$  alone unlike in the previous theorem.

We note also that here, the first condition in (13) can be interpreted as given the past and the present value of  $y$ , the dropout probabilities do not depend on  $\theta$ . In the context of continuous time survival analysis, Kalbfleisch and Prentice (pg. 121, 1980) call an analogous condition as *noninformative censoring*. Moreover, it is difficult to imagine both conditions of (13) holding except when the random dropout condition holds. The distribution of  $Y_k$  is characterized by  $\theta$  and if both  $f_{D_k|Y_k}(1|\text{past}_k, Y_k; \psi)$  and  $f_{D_k}(1|\text{past}_k; \psi)$  do not depend on  $\theta$ , then  $D_k$  must practically be independent of  $Y_k$  given the past.

Laird (1988) comments that the asymptotic variance of the estimator should be estimated by the observed information. The observed information matrix is given by the average over subjects of minus the second derivative of the log likelihood and the expected information is the expected value of this average. Under the conditions of Theorem 1 or 2, the observed information matrix depends only on  $L^C(\theta)$  and is thus a consistent estimator of the asymptotic variance. The expected information must be calculated relative to the density in (11). Therefore, the conditions of Theorem 1 or 2 are not sufficient for the consistency of the expected information. If we calculate the expectation relative to the density in (10) (that is, if we ignore the dropout) we are requiring that certain marginal first and second moments be preserved. This necessitates stronger conditions for ignorability as discussed below.

Suppose we wish to make inference about parameters in the mean  $E[\mathbf{Y}|\mathbf{X}] = \mu$  via a generalized estimating equation (GEE) as in Liang and Zeger (1986). Denote the subset of  $\theta$  contained in the mean by  $\beta$ . For a known function,  $g$ , the  $j$ th component of  $\mu$  is  $\mu_j = g(X_j, \beta)$ . Let  $V(\alpha)$  be a known function of  $\mu$  and a nuisance parameter  $\alpha$ , which is a guess at the covariance matrix of  $\mathbf{Y}$  given  $\mathbf{X}$  and let  $\tilde{V}$  be its inverse. If we have ideal data, the estimating function for  $\beta$  is then the sum over subjects of

$$\sum_{i=1}^K \sum_{j=1}^K \frac{\partial \mu_i}{\partial \beta} \tilde{V}_{ij}(\alpha) (Y_j - \mu_j).$$

However with incomplete data we may *ignore the dropout* and only use the available responses, that is we would use, the sum over subjects of

$$\sum_{i=1}^K \sum_{j=1}^K D_i \frac{\partial \mu_i}{\partial \beta} \tilde{V}_{ij}(\alpha) (Y_j - \mu_j) D_j \tag{14}$$

as the estimating function. If the above equation is still an unbiased estimating equation, i.e. it has mean zero, then we say that the *dropout is ignorable*.

We now give the conditions for the ignorability of the dropout under this type of inference.

**Theorem 3** *The dropout is ignorable under the preservation-of-marginal-moments type of inference if,*

1. *the independent dropout II condition holds for  $k = 1, \dots, K$ , and;*
- 2.

$$D_k \text{ is independent of } (Y_1, \dots, Y_{k-1}) \text{ given } D \geq k \quad (15)$$

*for  $k = 1, \dots, K$ .*

Note that (14) will have expectation 0 if for  $i \geq j$ ,  $E(Y_j|D > i) = \mu_j$ . Now, we have that  $E(Y_j|D > i) = E\{E(Y_j|\text{past}_j, D > i)|D > i\}$ . By the conditions of the above theorem,  $E(Y_j|\text{past}_j, D > i) = E(Y_j|Y_1, \dots, Y_{j-1})$  and that  $(Y_1, \dots, Y_{j-1})$  is independent of  $\{D > i\}$ . These give us the result.

Additionally under the ignorability conditions of the above theorem, the sandwich estimator (Liang and Zeger, 1986) based only on each subjects' observations up to his/her time  $D$  is a consistent estimator of the variance of the estimator of  $\theta$ .

Liang and Zeger (1986) require the data to satisfy a frequentist version of *missing completely at random* (MCAR) for the consistency of the estimators obtained from generalized estimating equations in the presence of missing data. This is the same as assuming that  $D$  and  $Y$  are independent which is stronger than the above conditions. The Bayesian version of *missing completely at random* (MCAR) is described by Little and Rubin (1987) and is satisfied when the missing data is Bayesian MAR and the observed data is *observed at random*. Rubin (1976) defines this condition as the probability that the dropout occurs at the observed time of dropout  $d$  is not functionally related to  $Y$ . As he notes this implies independence of  $Y$  and  $D$  only if we assume the MCAR condition for all possible values of  $D$  and not only the observed  $D = d$  (resulting in the frequentist version).

Robins et al. (1995) use weighted estimating equations where the weights are based on the response probabilities to get consistent estimators under the less restrictive Equation (8) (which is equivalent to frequentist MAR in our formulation). They

make the additional assumption that the response probabilities given the past are known up to a vector of unknown parameters, that is, the response probabilities can be expressed as a known function of unknown parameters, covariates not included in the model and past observations.

Suppose we wish to model the conditional mean  $\mu_j$  of  $Y_j$  given the past values  $(\mathbf{X}, Y_1, \dots, Y_{j-1})$  via a projected partial score as in Murphy and Li (pg. 402, 1995). As before, denote the subset of  $\theta$  contained in the conditional mean by  $\beta$ . For a known function,  $g$ ,  $\mu_j = g(\mathbf{X}, Y_1, \dots, Y_{j-1}; \beta)$ . Let  $V(\alpha)_j$  be a known function of  $\mu$  and a nuisance parameter  $\alpha$ , which is a guess at the covariance of  $Y_j$  given  $\mathbf{X}, Y_1, \dots, Y_{j-1}$ . If we have ideal data, the estimating function for  $\beta$  is then the sum over subjects of

$$\sum_{j=1}^K \frac{\partial \mu_j}{\partial \beta} V_j^{-1}(\alpha)(Y_j - \mu_j).$$

However with incomplete data we may *ignore the dropout* and only use the available responses, that is we would use, the sum over subjects of

$$\sum_{j=1}^K \frac{\partial \mu_j}{\partial \beta} V_j^{-1}(\alpha)(Y_j - \mu_j)D_j \tag{16}$$

as the estimating function. If the above equation is still an unbiased estimating equation, i.e. it has mean zero, then we say that the *dropout is ignorable*.

Equation (16) has expectation 0 if  $E[Y_j | \text{past}_j, D_j = 1] = \mu_j$ . Thus, we have the ignorability condition for this type of inference in the following theorem.

**Theorem 4** *The dropout is ignorable under the preservation-of-conditional-moments type of inference if the independent dropout II condition (5) holds for  $k = 1, \dots, K$ .*

Note that In Murphy and Li (1995)'s example involving longitudinal data subject to dropout, the independent dropout II condition is satisfied. Additionally under the above ignorability conditions, their sandwich formula for the estimator of the variance of the estimator of  $\beta$  also remains valid when applied only to each subjects' observations up to his/her time  $D$ .

## 5. DISCUSSION

We have shown that the definition of ignorable dropout depends on the population of inference and the statistical methodology used. Since one sample can be used to make inference about more than one population, it is important to carefully define the population when dealing with premature departures from a longitudinal study. Not all subjects who leave the study prematurely are dropouts – the population of inference dictates whether a premature departure is a dropout or an end of a complete observation.

Ignorability conditions depend on the type of inference. Rubin (1976) considers ignorability for Bayesian inference. Since a Bayesian analysis is conditional on the collected sample, Rubin’s ignorability conditions concern only the collected sample. We have discussed three methods of frequentist inference. Since all three methods depend on large sample theory for the construction of confidence intervals and hypothesis tests, our ignorability conditions must hold for all possible samples collected from the population.

Our ignorability conditions appear weakest when interest is in the parameters of the conditional mean of  $Y_j|\mathbf{X}, Y_1, \dots, Y_{j-1}$ . However as we mention, the same ignorability condition, independent dropout II, is sufficient for a (partial) likelihood analysis. That is if (12) does not hold but independent dropout II holds, then the practitioner, in ignoring the dropout, is using a partial likelihood analysis rather than a full likelihood analysis.

## ACKNOWLEDGEMENT

This work was supported in part by the National Science Foundation and the National Institute on Aging.

## Reference

- ANDERSEN, P., BORGAN, O., GILL, R., & KIEDING, N. (1988). Censoring, truncation and filtering in statistical models based on counting processes. *Contemporary Mathematics*, 80, 19–60.
- COX, D. (1975). Partial likelihood. *Biometrika*, 62(2), 269–76.
- DE GRUTTOLA, V., & TU, X. M. (1994). Modelling Progression of CD4-Lymphocyte Count and Its Relationship to Survival Time. *Biometrics*, 50, 1003–1014.
- DIGGLE, P., & KENWARD, M. G. (1994). Informative dropout in longitudinal analysis. *Applied Statistics*, 43(1), 49–93.
- DIGGLE, P. J. (1989). Testing for Random Dropouts in Repeated Measurement Data. *Biometrics*, 45, 1255–1258.
- Economic Behavior Program, Survey Research Center, Institute for Social Research, The University of Michigan, Ann Arbor, Michigan (1984). *User Guide to the Panel Study of Income Dynamics*.
- FISHER, L., DIXON, D., HERSON, J., FRANKOWSKI, R., HEARRON, M., & PEACE, K. (1990). Intention to treat in clinical trials. In Peace, K. (Ed.), *Statistical Issues in Drug Research and Development*. Marcel Dekker.
- HEYTING, A., TOLBOOM, J., & ESSERS, J. (1992). Statistical Handling of Dropouts in Longitudinal Clinical Trials. *Statistics in Medicine*, 11, 2043–2061.
- HILL, M. (1992). *The Panel Study of Income Dynamics: A User's Guide*. Sage Publications, Inc.
- HOGAN, J., & LAIRD, N. (1995). Intention-to-treat analyses for incomplete repeated measures which depend upon event times. *to appear in Biometrics*.

- HOGAN, J., & LAIRD, N. (1996). Model-based Approaches to Analyzing Incomplete Longitudinal and Failure Time Data. *to appear in Statistics in Medicine*.
- HOOVER, D., MUNOZ, A., CAREY, V., TAYLOR, J., VANRADEN, M., CHMIEL, J., & L.KINGSLEY (1993). Using events from dropouts in nonparametric survival function estimation with application to incubation of AIDS. *Journal of the American Statistical Association*, 88(421), 37–43.
- KALBFLEISCH, J., & PRENTICE, R. (1980). *The Statistical Analysis of Failure Time Data*. John Wiley and Sons, Inc.
- KASLOW, R. A., OSTROW, D., DETELS, R., PHAIR, J., POLK, B., & RINALDO, C. (1987). The Multicenter AIDS Cohort Study: Rationale, Organization, and Selected Characteristics of the Participants. *American Journal of Epidemiology*, 126(2), 310–318.
- LAIRD, N. M. (1988). Missing Data in Longitudinal Studies. *Statistics in Medicine*, 7, 305–315.
- LIANG, K., & ZEGER, S. (1986). Longitudinal data analysis using gnerealized linear models. *Biometrika*, 73(1), 13–22.
- LITTLE, R. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90(431), 1112–1121.
- LITTLE, R., & RUBIN, D. (1987). *Statistical Analysis with Missing Data*. John Wiley and Sons, Inc.
- MURPHY, S., & LI, B. (1995). Projected partial likelihood and its application to longitudinal data. *Biometrika*, 82(2), 399–406.
- ROBINS, J., ROTNITZKY, A., & ZHAO, L. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90(429), 106–21.

- RUBIN, D. B. (1976). Inference and Missing Data. *Biometrika*, 63(3), 581–92.
- SCHLUCTER, M. D. (1992). Methods for the Analysis of Informatively Censored Longitudinal Data. *Statistics in Medicine*, 11, 1861–1870.
- SCHWARTZ, D., & J.LELLOUCH (1967). Explanatory and pragmatic attitudes in therapeutic trials. *Journal of Chronic Diseases*, 20, 637–48.
- Survey Research Center, Institute for Social Research, The University of Michigan, Ann Arbor, Michigan (1986). *A Panel Study of Income Dynamics: Procedures and Tape Codes, 1984 Interviewing Year, Wave XVII, A Supplement*.
- WU, M. C., & CARROLL, R. J. (1988). Estimation and Comparison of Changes in the Presence of Informative Right Censoring by Modeling the Censoring Process. *Biometrics*, 44, 175–188.
- WU, M., & BAILEY, K. (1989). Estimation and Comparison of Changes in the Presence of Informative Right Censoring: Conditional Linear Model. *Biometrics*, 45, 939–955.