Journal of Biopharmaceutical Statistics, 10(4), 503–525 (2000)

DROPOUTS IN LONGITUDINAL STUDIES: DEFINITIONS AND MODELS

J. K. Lindsey

Department of Biostatistics
Limburgs University
3590 Diepenbeek, Belgium

Key words. AIC; Dropouts; Longitudinal study; Missing data; Proportional hazards; Survival distribution; Time-varying covariate; Weibull distribution

Abstract

The widely used distinction of Little and Rubin (1) about types of randomness for missing data presents difficulties in its application to dropouts in longitudinal repeated measurement studies. In its place, a new typology of randomness for dropouts is proposed that relies on using a survival model for the dropout process.

In terms of a stochastic process, dropping out is a change of state. Then, the longitudinal measures and dropout processes can be modeled simultaneously, each conditional on the complete previous history of both repeated measures and states. In this context, Poisson regression is used to fit various proportional hazards models, some of which are new, to the dropout process using the longitudinal measurements responses as time-varying covariates.

As examples of longitudinal measurement studies displaying nonrandom dropout processes, a dental study of testosterone production in rats and clinical trials for treatment of gallstones and of depression are analyzed.

1. Introduction

1.1 Role of Dropouts

The problem of missing observations is important in any study, but it is particularly acute when repeated measurements are taken on individuals. The missing values may occur in the response variable or in the explanatory variables or both. In sample surveys, for example in epidemiology, nonresponse in the explanatory variables is usually of special concern because of the cumulative effect in multiple regression models. In panel studies, certain individuals may not be contactable at some of the points in time (waves) so that all variables are missing there. In experimental trials, baseline explanatory variables are generally recorded, along with the treatment to which each individual is randomized. In such longitudinal studies, two different missingness situations among the repeated responses can be distinguished: an occasional response value may be missing, or the individual may drop out so that all response values are missing after some point in time. It is this last situation that will be examined here.

In almost all circumstances, dropouts in longitudinal studies are not simply a special case of missing values. In contrast to the occasional missing value, dropping out of a properly planned clinical trial generally requires a major decision to leave the study completely. If dropping out can be assumed to have occurred randomly, in the sense to be discussed below, then standard analyses can be applied that simply ignore the subsequent missing values, although with the attendant loss of information in certain cases. Many methods are now widely available for such unbalanced data (Lindsey (2)). However, in studies involving living beings, it may generally safely be assumed that observations are rarely randomly missing in the sense to be used in this paper (as opposed, for example, to test tubes accidentally being dropped) and even less that dropping out occurs randomly.

It is often argued that ‘‘scientific’’ interest should focus on the longitudinal profile averaged over dropout patterns, apparently in an attempt to determine what might have happened if there had been no dropouts. The fundamental thesis of this paper is that such an approach mixes up, in an uninterpretable way, two processes that are occurring in the study: the generation of the longitudinal series of responses and the reactions to the conditions of the study that produce dropouts. Such marginal results, averaged over dropout patterns, will yield misleading conclusions if subsequently applied in any real situation where dropouts are possible.

1.2 Current Approaches

Much has been made about the distinctions among nonrandom, random (MAR), and completely random (MCAR) missing values (Little and Rubin (1)). However, these ideas are quite distinct from those that will be used here because they involve

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

the counterfactual principle of unrecorded values. They implicitly assume that the ''as if'' situation of no dropouts is of primary importance when usually it is an unrealistic hypothetical possibility of no real interest. These ideas are difficult, if not impossible, to apply empirically because, by definition, it is impossible to determine if the reason that a value is missing depends on the values of observations that were not recorded.

Many papers have been written in recent years about missing values. No attempt will be made to review the literature here. A considerable number of these papers have been specifically concerned with the dropout problem in longitudinal studies. Some seek to determine if the dropouts are random in the Little and Rubin sense (for example, Diggle (3); Ridout (4)). Others construct models for the observed responses in the presence of missing data (for example, Gould (5); Murray and Findlay (6); Heyting, Tolboom, and Essers (7); Kenward, Lesaffre, and Molenberghs (8)), while some even attempt the empirically impossible task (because no information is available) of modeling dependence of present dropout on hypothetical unobserved future values (for example, Diggle and Kenward (9); Lesaffre, Molenberghs, and Dewulf (10); Molenberghs, Kenward, and Lesaffre (11)). For a review, see Little (12).

A number of authors have developed models for the simultaneous analysis of the longitudinal measurements and the time to dropout, by conditioning either dropout on the repeated responses in a selection model, (for example, Woodbury, Manton, and Stallard (13); Manton, Woodbury, and Stallard (14); Wu and Carroll (15); Schluchter (16); De Gruttola and Tu (17); Tsiatis, De Gruttola and Wulfsohn (18); Martinussen and Keiding (19)), or the repeated measures on dropping out in a pattern mixture model (for example, Wu and Bailey (20,21); Mori, Woodworth, and Woolson (22); Pawitan and Self (23); Hogan and Laird (24)); for an overview, see Hogan and Laird (25). Another approach, closely related to the latter ones, is to model the numbers of recorded observations (Mori, Woolson, and Woodworth (26)) instead of time. Note that making the distribution of the observed longitudinal measures conditional on time to dropout, as in a pattern mixture model, means conditioning what is happening in the present on a future event because dropout occurs after the last measurement is made.

Thus, many current approaches to dropouts attempt to remedy defects in study design (unrecorded reasons for dropout) by making unverifiable modeling assumptions where, as Fisher (27) stated, only a post mortem is feasible. The solution is to collect information on reasons for dropouts, not to make unverifiable assumptions about dependence on unobserved events after dropout.

1.3 Dropout as a Survival Process

Here, I shall use survival models to study the dropout process simultaneously with a model for the longitudinal measurements process, both conditional on the

previous history of each subject. This procedure can efficiently use the available information to determine whether the dropout process is random or not, in a way to be defined below. But, in most cases, it can also provide important information that can, in its own right, be useful in a complete analysis of the data.

For example, if a study protocol specifies that patients be withdrawn when their blood pressure exceeds a certain level or that machines must be readjusted if the product leaves certain tolerance limits, then the only model of the longitudinal responses that makes sense is one that conditions on not crossing the threshold. But, at the same time, a complementary model for the risk of reaching that threshold must form an essential part of the analysis. The same reasoning will most often still hold when the causes of dropping out are not so clearly defined at the beginning of a study.

In the context of short series of longitudinal observations, as in panel data, Taris (28,29) uses Markov chains to determine if dropouts are random. He considers dropouts to be nonrandom if the Markov chain is not stationary and identical for all subgroups of individuals relevant to the question at hand. In a similar way, we can model dropouts using survival models to study randomness, checking if the risk of dropout changes over time and if it depends on the previous history of the individuals.

The goal of modeling dropouts, and the interpretation of the results, will depend on the type of longitudinal study involved. Here, I shall consider as examples a dental study of testosterone production in rats and two multicenter clinical trials, one involving repeated counts of nausea in patients with gallstones and the other for treatment of depression.

In summary, the approach taken here is to model, as completely and adequately as possible, those responses that have been observed, as they depend on the observed previous history of subjects, rather than to postulate hypothetical and uncheckable models for what might have been observed if circumstances had been otherwise.

2. Types of Dropouts

2.1 Basic Principles

2.1.1 Standard Definitions

The widely used Little and Rubin (1) distinctions among nonrandom, random, and completely random missingness have often been distorted or misinterpreted in the literature. They are founded on three basic ideas: (1) data are missing at random if, given the observed data, the probability of the observed missingness pattern does not depend on the values of the unobserved data; (2) data are observed at random if, for every possible value of the missing data, the probability of the observed missingness pattern, given the unobserved data, does not depend on the

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

values of the observed data; and (3) the parameters for the data and for the missingness models for the complete data are distinct if knowledge of one does not place constraints on the other's values (Rubin (30)). In spite of their names, these definitions do not correspond to any commonly accepted definition of randomness but, instead, describe *independence* between the response mechanism of interest and the missing data mechanism. Thus, the Little and Rubin (1) distinctions should rather be called dependent, independent, and completely independent missingness.

Such ideas originally arose in the context of cross-sectional sample surveys. In such a context, they may often be useful to model the missing responses of people whom it was not possible to contact as compared to those who refused to participate. They are less clearly appropriate in other cases, such as for refusals to specific questions in such surveys; here, for example, for questions of opinion, the only possible response is ''missing,'' whereas, for income, a relevant objective value does exist. In these cases of refusal, assuming any other form for the ''unobserved'' data, including imputing what would have been the response, can bias the results. Studies of sample survey interviewing procedures have shown that such individuals are usually very different from those who agree to reply. Such people generally have good reasons for not responding. However, no empirical information is generally available about these differences.

2.1.2 Dropouts

In a longitudinal measures study, a *dropout*, in the terminology of stochastic processes, corresponds to a change of state of the subject. In cases of interest, this generally occurs because of the previous history of the subject within the study (for example, lack of patient improvement or side effects) or because a response is no longer possible (for example, refusal or death). More occasionally, withdrawal may be because subsequent responses will no longer be relevant to the phenomenon under study due to changing circumstances (for example, patient recovery, moving away, or contracting an unrelated disease). Averaging over profiles with such different dropout patterns (and reasons) can have little meaning in such circumstances. This will be especially true in a clinical trial where noncompliance and dropout are strictly controlled and hence not representative of any subsequent general use of a medication.

It is difficult to conceive of a situation in which recording would stop, knowingly or unknowingly, simply because the next response is expected to be extreme (but still meaningful) without basing the decision on the subject's previous history. Thus, in most situations, it does not make sense to require the dropout and longitudinal measures mechanisms to be independent: recording each repeated measure will be dependent on not previously having dropped out while the risk of dropping out can only depend empirically on the (series of) repeated measures, and covariates, already recorded. Present events do not depend on hypothetical future events except through prediction from knowledge of previous events.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Dropouts in longitudinal measurements studies have often been associated with censoring in survival data. However, the situations are quite different. In the latter case, it is important that the censoring mechanism not depend on the failure mechanism, for example by subjects being removed from the study because of a high risk of failure (Kalbfleisch and Prentice (31), Ch. 5). The principal case in which this can be a real problem in longitudinal repeated measures studies is if only the final response is to be analyzed (Heyting et al. (7)) instead of modeling the complete longitudinal series. Of course, due to dependence among responses on an individual, earlier responses will contain information about this final response that is lost if they are ignored.

2.2 Problems with Standard Definitions

In this light, Rubin's (30) definitions pose several major problems in the context of dropouts in longitudinal data, including the following:

1. In many situations where subjects drop out, they are no longer in a position to provide a response so that the only possible value of the missing data is ''missing.''
2. Even in those cases where a further response would be possible after dropout, stating that the probability of dropout depends on its hypothetical value means conditioning on the future because dropping out occurs before that response would have been recorded. If a patient is feeling too ill to continue visits, the model should condition on this information, not on the hypothetical value that otherwise would have been recorded.

In almost all cases of interest, unrecorded responses after dropout (where they are possible) will, by definition, be qualitatively different from those of subjects who have not dropped out and cannot be directly compared with them (except in the rare case of random dropout, in the sense used here). The subjects have changed state so that the process generating the longitudinal measurements (observed or not) is different (although this cannot be checked empirically unless subsequent observations are available).

Some have argued that conditioning on the unobserved response values is similar to conditioning on an unobserved random effect. However, the response could have been observed whereas the random effect never can be. Indeed, the random effect disappears from the model actually used to construct the likelihood function, its role being to generate dependencies among the observed responses on a subject (Lindsey (2, pp. 55–61)). In the same way, a model claiming to condition on the unobserved response makes strong implicit assumptions in order to produce a prediction that must, in fact, be implicitly conditional on the previous observed values. In empirical implementation, any such model, thus, must actually

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

assume MAR. It would seem preferable to condition directly on the observables rather than indirectly in this obscure way.

For frequentist inference, the MAR dropout process creates major problems. It is widely stated that frequentist likelihood inference is possible in such a case; it has even been claimed that standard errors obtained from the observed information matrix are valid (for example, Kenward and Molenberghs (32,33)). However, the dropout process is acting as a stopping rule that depends on previous observations, as in sequential designs. The dropout process determines how much information is collected, and this depends on what information was already collected. As for sequential designs, such a stopping rule cannot be ignored in making frequentist inferences. Such problems do not arise in direct likelihood and Bayesian inference.

Thus, we must accept the fact that, in most cases where dropouts are present, the only interpretable model will be one involving the response profiles over time conditional on not (yet) dropping out. In this sense, only modeling the complete *observed* series of longitudinal responses prior to dropout is ''unbiased,'' but it needs to be complemented by a model for dropping out. On the other hand, forcing subjects to remain in a study when they want to or should drop out will bias the results, as will theoretical attempts to guess (impute) what might have been the response after dropout, because the dropouts are different than those who would stay in the study anyway.

2.3 Defining the Dropout Process

Let us then consider how to model the dropout process using an accepted definition of randomness (not that of Rubin (30)) and without involving the hypothetical responses after dropout. First note that all subjects who do not drop out are censored in terms of the dropout process. This is uninformative because, in standard survival analysis terminology, it is Type I censoring, determined by the study design.

2.3.1 Factors Affecting Dropout

In any longitudinal study, risk of dropout can be modeled as depending on a number of observable factors. The first to consider is time. If risk, for all subjects, can be shown not to vary in time, then one may postulate a Poisson process, implying that dropping out is random over the period of the study. If it varies over time in the same way for all subjects (for example, because of aging, unwillingness to continue participation, and so on), dropouts are not random but the dropout process is ignorable because it is not connected to the phenomenon under study, whether the longitudinal response, treatment, or relevant covariates.

If the risk of dropout is constant over time within some subgroups determined at the beginning of the study (treatment and/or baseline covariates), but

different among such subgroups, then it can be taken to be random within these subgroups. (For example, those with complete series are a random subsample from the sample subgroup; see Heyting et al. (7).) However, if a random sample from some population was chosen at the beginning of the study, it is becoming less and less representative over time because of the differential rates of dropout in the different subgroups. On the other hand, in an experimental trial where subjects are randomized to treatments, differential random dropout among treatments may not cause too much concern, at least for a treatment comparison at some final endpoint. Thus, this type of dropout process is not random for the study as a whole because of the dependence on these subgroups. In almost any longitudinal context, this dropping out is not ignorable.

In summary, things will generally not be simple. In addition to time, the risk of dropout may depend on treatment, baseline covariates, time-varying covariates, and the values of the longitudinal response variable already recorded. In all of these cases, the dropout process is nonrandom. In randomized experimental trials, if the risk of dropping out depends only on treatment, then causal conclusions can be made about it. If it depends on time-varying covariates or on previously recorded responses, conclusions may be more difficult to draw because these will generally be intermediate, unrandomized variables, depending on previous values, on baseline covariates, or on treatment.

2.3.2 Types of Dropout Mechanisms

We therefore find three distinct dropout processes:

1. If risk can adequately be described by the same homogeneous Poisson process for all subjects, dropout is *random*.
2. If risk varies over time, or depends on irrelevant factors external to the study, in the same way for all subjects, the process is *ignorably nonrandom*.
3. If risk depends on any of the variables relevant to the process under study, including those used in the model for the longitudinal measures process and any specifically collected as reasons for dropping out, dropout is *nonignorably nonrandom*.

Naturally, extreme care must be taken when concluding that the dropout process is random or ignorable because the appropriate dependence on the available variables may not have been discovered.

If the dropout process is nonignorable (nonrandom), the subjects change state and (possible) future unrecorded longitudinal responses will generally be irrelevant to the repeated measures process before change of state (dropout). In the presence of nonignorable dropouts, asking what outcomes would have been ob-

served if all subjects had completed the study will have no meaning in most situations.

This contrasts with the intention-to-treat situation where treatment differences measure the social phenomenon of assigning medication, not the biological process of taking it. After dropout, we do not have observations over which to average.

3. Modeling the Dropout Process

3.1 Risk of Dropout

3.1.1 Proportional Hazards Models

To model the dropout process using these definitions, we require a flexible procedure for survival data that can easily accommodate time-varying covariates. Lindsey (34) has shown how parametric proportional hazards models for failure time data can simply be fitted as log linear models, that is, by Poisson regression, a procedure that is now available in most major statistical packages such as Genstat, GLIM, Lisp-Stat, R, S-Plus, SAS, or Stata. In our context, the basic model has the form

$$\log[\lambda(t;\,\boldsymbol{\theta})] = \beta_0 + \sum_i \beta_i g_i(t) + \sum_j \gamma_j x_j + \sum_k \eta_k z_{kt}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta})$; $\lambda(\cdot)$ is the risk or hazard function for dropping out, x_j are the baseline covariates and treatments; z_{kt} are time-varying covariates and (one or more previous) longitudinal measures responses; and t indexes time. Here, each $g_i(t)$ is some completely known function of time. Common choices include t, t^2, $1/t$, $1/t^2$, $\log(t)$, and $\log^2(t)$. Because the series of observations on each individual is usually short, at most two will generally be necessary in a given model. As well as monotone hazard functions, both unimodal and bathtub shapes are possible in this family. Interactions between time and the other variables can also be included; without them, the models are in the family of proportional hazards or multiplicative intensities.

Several members of this family are well known:

1. When there is no function of time, we have the exponential distribution, that is, the homogeneous Poisson process.
2. when the only function in time is
   $g_1(t) = \log(t)$, we have the Weibull distribution,
   $g_1(t) = t$, we have the extreme value distribution,
   A factor variable taking a different level for each time point, we have the Cox (35) model.

For no other special case of the model is a simple density available (Lindsey (36)).

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

3.1.2 Integrated Hazard

In constructing models, we require the following integrated hazard (I ignore the covariates for clarity):

$$\Lambda(t;\,\beta) = \int_0^t \lambda(\mu;\,\beta)du$$

$$= \int_0^t \exp\left[\beta_0 + \sum_i \beta_i g_i(u)\right]du$$

For other functions of t than those listed above, this is intractable. However, for some of them, well-known numerical procedures are available. Thus, for example, for $g_1(t) = t$ and $g_2(t) = t^2$, we obtain the normal integral; for $g_1(t) = \log(t)$ and $g_2(t) = \log^2(t)$, the log normal integral; for $g_1(t) = t$ and $g_2(t) = \log(t)$, the gamma integral; and for $g_1(t) = t$ and $g_2(t) = 1/t$, the inverse Gaussian integral. However, note that these are integrated hazards, not cumulative probabilities. When this integral, from zero to infinity, is finite, the survival distribution is ''defective'' in that there is a finite probability of never having the event. In our context, this means that some individuals will never drop out, a useful characteristic of these models.

Consider in more detail, for example, the model where $g_1(t) = t$ and $g_2(t) = t^2$, with an integral that would correspond to a normal density. Let

$$\lambda(t;\,\beta_1,\,\beta_2) = \exp[\beta_1 t + \beta_2 t^2 + \beta_1/(4\beta_2) + \log(2\beta_2/\pi)/2]$$

the canonical form of the normal density [with $\beta_0 = \beta_1/(4\beta_2) + \log(2\beta_2/\pi)/2$], but here taken as a hazard or risk function. Then, the integrated hazard is the integral of this and the corresponding survivor function is

$$S(t;\,\beta_1,\,\beta_2) = \exp\left[-\int_0^t \lambda(u;\,\beta_1,\,\beta_2)du\right]$$

Notice that, because the integral to infinity is finite, there is nonzero probability of surviving forever, that is, not dropping out. The corresponding density is

$$f(t;\,\beta_1,\,\beta_2) = \lambda(t;\,\beta_1,\,\beta_2)\exp\left[-\int_0^t \lambda(u;\,\beta_1,\,\beta_2)du\right]$$

This belongs to the class of nonstandard parametric proportional hazards or multiplicative intensities models that will be used here.

3.2 Likelihood Function

3.2.1 Construction

With this approach, the model for dropouts is simple to implement using generalized linear models in any of the software mentioned above. I assume that dropping

out occurs sometime after the last recorded response, and prior to the time when the next response would have been recorded. Then, the dropout response for the Poisson regression is an indicator variable, say **d**, of change of state consisting of a series of 0's at all time points, sometime after when each longitudinal measure is recorded, except for that at dropping out, after the last recorded response, which is a 1. Subjects who do not drop out are uninformatively censored; they are handled automatically by a series of 0's with no 1 at the end. All of the variables in the longitudinal measures study, including the responses up to that time, form part of the previous history of a subject and can be used as explanatory variables, if necessary. If the times between observations are irregular, the logarithm of these intervals, Δ_t, is used as an offset. This will often be fairly crude because the exact time of dropping out may not be recorded; sometimes, even the timing of the first missing longitudinal measurement is unknown.

Now let t indicate the time of recording a longitudinal measurement. The log likelihood for this dropout model for one subject is (Lindsey (34))

$$\log[\mathbf{F}_D(\beta;\, \mathbf{d})] = \sum_{t=1}^{n} d_{t_+} \log[\lambda(t_+|\mathcal{F}_t;\, \beta)\Delta_t] - \sum_{t=1}^{n} \lambda(t_+|\mathcal{F}_t;\, \beta)\Delta_t$$

where n longitudinal measurements are recorded; d_{t_+} is the dropout indicator variable for change of state of the individual, with the $+$ indicating that it is recorded after the longitudinal measure at time t; $\mathcal{F}_t$ is the relevant history (covariates and repeated measures responses) up to, but not including, time t_+; and Δ_t is the interval between observations.

Then, the complete likelihood function for one subject based on the joint distribution of the longitudinal measurements, conditional on not having dropped out, that is not having changed state, and on the dropout process is

$$\begin{aligned}\mathbf{L}(\boldsymbol{\alpha},\, \boldsymbol{\beta};\, \mathbf{y},\, \mathbf{d}) &= \mathbf{L}_{RM}(\boldsymbol{\alpha};\, \mathbf{y}|\mathbf{d} = \mathbf{0})\mathbf{L}_D(\boldsymbol{\beta};\, \mathbf{d}) \\ &= \prod_{t=1}^{n} f(y_t|\mathcal{F}_{t-},\, d_{t-} = 0;\, \boldsymbol{\alpha})[\lambda(t_+|\mathcal{F}_t;\, \beta)\Delta_t]^{d_{t+}} \exp[-\lambda(t_+|\mathcal{F}_t;\, \beta)\Delta_t] \quad (1)\end{aligned}$$

where y_t is the observed longitudinal measures response, recorded before d_{t_+} occurs, with conditional density function, $f(\,)$, and parameters, $\boldsymbol{\alpha}$. Because of the conditioning, the complete likelihood function factors into the two separate likelihood functions, the repeated measurements likelihood, $\mathbf{L}_{RM}$, and the dropout likelihood, $\mathbf{L}_D$, so that, with direct likelihood and Bayesian methods, they can always be studied separately (unless, for some unusual reason, they have common parameters).

3.2.2 Interdependence Between Processes

As time goes by, the dropout process is imposing a selection on those individuals still able to supply repeated measurements. This implies that the form of the distri-

bution of longitudinal responses may be changing over time. Under certain very strict mathematical conditions (Woodbury et al. (13); Manton et al. (14); Martinussen and Keiding (19); Yashin and Manton (37)) on both processes in the model, the distribution of the longitudinal measure, $f(y_t|\mathcal{F}_{t-}, d_{t-} = 0; \boldsymbol{\alpha})$, at successive time points can be made to adjust automatically to this selection process. The greater flexibility of allowing the adjustment to take place through dependence on the previous history of each individual, $\mathcal{F}_{t-}$, seems preferable. Thus, any appropriate standard longitudinal repeated measures model can be used to analyze, conditionally, all observed responses up to dropout. These aspects of the complete model will not be considered further here. It is rather the second of the two factors in the likelihood of Eq. (1) that is of interest.

On the other hand, although the likelihood factors, the longitudinal measurement process, **Y**, and the dropout process, **D**, are, in general, not even conditionally independent so that their joint multivariate distribution does not factor into parts containing each separately. This contrasts with the simpler situation assumed in the literature on joint modeling mentioned in the introduction where either $f(\mathbf{y})f(\mathbf{d}|\mathbf{y})$, a selection model, or $f(\mathbf{y}|\mathbf{d})f(\mathbf{d})$, a pattern mixture model, is used. Thus, if the marginal distribution of the longitudinal responses is required here, it can only be obtained by integrating Eq. (1) over possible dropout times; this will usually be complicated because $\mathcal{F}_t$ will generally contain these responses, either in the dropout or the repeated measures model or both. This problem will also not be considered further here, the reason being that—contrary to frequent statements in the missing data literature—this marginal distribution generally represents the ''as if'' situation of no dropouts corresponding to no real phenomenon of interest.

Thus, in the complete analysis of any study, the two processes must both be considered. For example, in a clinical trial, if either the profile of longitudinal responses or the risk of dropout, or both, depend on the treatment, then, from the randomization, we can conclude that there is evidence of a causal treatment effect, although the importance of each must be suitably weighed. In longitudinal repeated measurements studies with dropouts, the conclusions must necessarily be rather complex.

4. Examples

The inference criterion that I shall use for comparing the models in the examples below will be their ability to predict the observed data, that is how probable they make the complete observed set of dropout sequences, including the censored ones. In other words, models will be compared directly through their minimized -2 log likelihood (Lindsey (38,39)). When the numbers of parameters in models differ, they will be penalized by adding twice the number of estimated parameters,

the Akaike information criterion (AIC) (see Akaike (40); Lindsey and Jones (41)). Smaller values indicate more preferable models.

In the following examples, I have considered all of the functions of time mentioned above, but only present results for those that proved most adequate for the given data set. For simplicity, I only look at dependence of dropout risk on the immediately previous observed value of the longitudinal response. However, if risk depends on the rate of change of these responses, as it well may, then two or more previous responses would need to be included. Thus, I am almost certainly underestimating the possibility of nonignorable dropouts.

4.1 Testosterone Production in Rats

Dentists interested in the therapeutic use of hormones set up an experiment to investigate the effect of inhibition of the production of testosterone on the craniofacial growth of male Wistar rats (Verbeke and Lesaffre (42)). The rats were randomized either to control or to low or high dose of decapeptyl, an inhibitor of testosterone production, with, respectively, 15, 18, and 17 in each group. Treatment was started when the rats were 45 days old, with 7 measurements being made every 10 days, starting when they were 50 days old. Each rat was anesthetized in order to X-ray its skull and measurements were made between various pairs of points on the resulting pictures. Unfortunately, many rats did not survive anesthesia, as can be seen in Table 1 (we do not know how many rats died at the last measurement).

In such an experimental context, one would not expect either the treatment or previous measurements to affect dropout. As well, the dentists were convinced that the risk of dying should not change over time. The first expectations were confirmed by the data, whereas the last was not: risk varied over time, in the same way for all treatment groups and independently of previous measurements, as can be seen in Table 2. Although the highest proportion of rats died in the control and the lowest in the high dose group, there is no indication of treatment influenc-

Table 1. Numbers of Rats Dying at Different Ages in the Three Treatment Groups

Age	Control	Low dose	High dose
50	2	1	1
60	0	2	1
70	3	0	2
80	3	3	3
90	3	2	0
100	0	2	0

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Table 2. AICs for Several Models for Dropouts in the Rat Experiment

Model	AIC	Parameters
Null	90.5	1
Treatment	91.2	3
Response	91.2	2
Time	91.5	2
Time + Time^2	89.0	3
Time + Time^2 + Treatment	89.8	5
(Time + Time^2) * Treatment	92.3	9
Time + Time^2 + Response	90.0	4
(Time + Time^2) * Response	90.1	6

ing mortality. Thus, we may conclude that the dropout process is nonrandom but that it is ignorable because it does not depend on the process under study.

For these data, the risk functions have the form of a normal density. The risk curve (for all treatment groups) is plotted in Figure 1. The nonrandomness might possibly arise from the weaker rats being eliminated over the first four or five anesthetics or by a learning process whereby, over time, the investigators

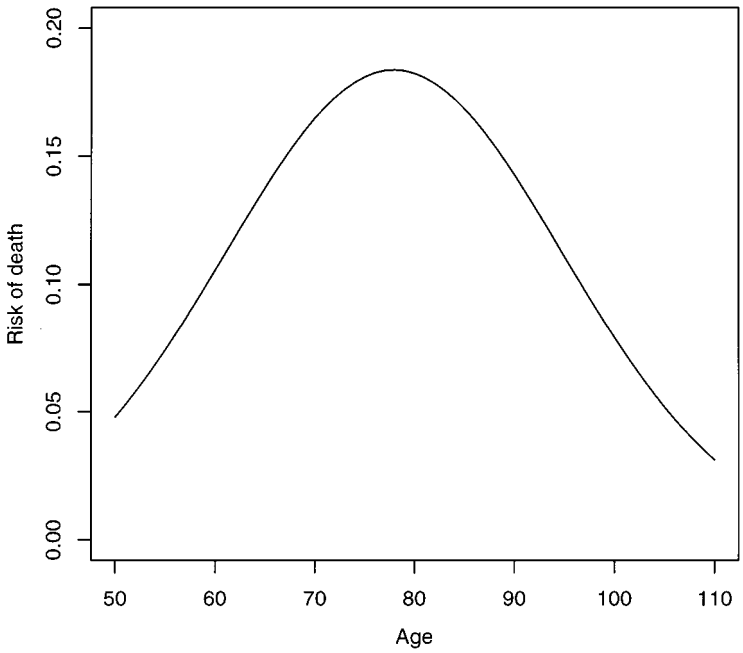

Figure 1. Risk curve for dropout from the rat experiment.

improved their techniques for handling the rats. In their calculation of efficient sample designs, Verbeke and Lesaffre (42) ignore this changing risk of dropout.

4.2 Nausea with Gallstones

Thall and Lachin (43) describe a 10-year, multicenter, double-blinded clinical trial to study the use of natural bile acid, chenodeoxycholic acid, to dissolve cholesterol gallstones. A total of 916 patients were randomized to be treated for up to 2 years each with either high or low dose or placebo. Patients were scheduled to return for clinic visits at 1, 2, 3, 6, 9, and 12 months, at which times counts of episodes of nausea were recorded. However, the patients actually visited at rather erratic times; these are used below. Only those observations for the first year (or more exactly, including the first visit after 58 weeks) for patients with floating gallstones in the high dose and placebo groups are given by Thall and Lachin, for a total of 111 patients (two in the high dose group dropped out before the first visit). Several reasons for exit from the study are given: dropout, withdrawal, hepatoxicity, and cholecystectomy, for a total of 16 patients (not including the two who dropped out early), one-half in each group. Because of the small numbers of each, it is not possible to consider these reasons separately in the analysis.

The results of fitting the dropout models to these data are shown in Table 3. There is little or no indication of a (constant over time) difference in the risk of dropout with treatment or response (count of nausea spells). However, it does depend on time (and its square). Because of the irregularity of the visits, it is not feasible to fit a Cox model to so few events. Introducing an interaction between time and treatment improves the model, whereas that with response does not. Again, the risk functions have the form of a normal density.

In the placebo group, the risk of dropout is high at the beginning, later reducing to about zero; on the other hand, in the high dose group, it shows a high level only between about weeks 15 and 40. The curves are plotted in Figure 2 where the normal shapes are clear. (The risk does not depend on previous response.)

Table 3. AICs for Several Models for Dropouts in the Gallstone Clinical Trial

Model	AIC	Parameters
Null	136.8	1
Treatment	135.5	2
Response	138.4	2
Time	124.2	2
Time + Time^2	121.5	3
(Time + Time^2) * Treatment	115.5	6
(Time + Time^2) * Response	121.9	6

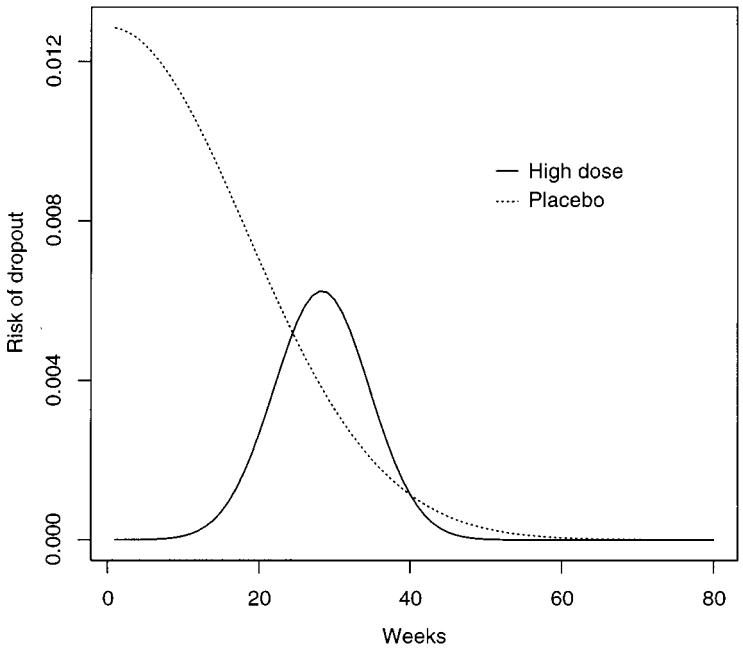

Figure 2. Risk curves for dropout in the two treatment groups for the gallstone trial.

Because of randomization of treatments, the difference between the curves can be interpreted causally. Note, however, that a satisfactory model would require a distinction among the reasons for dropping out; here this is not really possible because of the small number of patients leaving for each reason.

4.3 Antidepressant

Several large multicenter longitudinal clinical trials were conducted to test a compound belonging to a new class of antidepressants (Heyting et al. (7)). These involved a randomized, double-blind comparison of three compounds, including placebo, in parallel groups of outpatients with primary depression. The study groups were compared with respect both to efficacy and to unwanted effects. In the trial considered here, with eight centers, the target number of patients per center was 60, equally divided over the 3 treatment groups. The protocol stipulated an initial washout period during which the suitability of the patients for inclusion in the study was to be assessed according to inclusion and exclusion criteria. The washout period was followed by a treatment period of 6 weeks, during which study medication was given on a flexible basis, according to the patient's response. In the first treatment week, the dosage of the study medication was to be increased

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Table 4. AICs for Several Models for Dropouts in the Antidepressant Clinical Trial

Model	AIC	Parameters
Null	543.6	1
Treatment	545.3	3
Response	533.3	2
Time	542.3	2
Time + Time^2	528.5	3
Time + Time^2 + Treatment	530.3	5
Time + Time^2 + Response	519.2	4
Time + Time^2 + Response + ConcMed	517.6	5
Time + Time^2 + Response + ConcMed + FamHist	516.8	6
Time + Time^2 + Response + ConcMed + FamHist + Centre	513.9	13

ConcMed-binary indicator whether receiving concurrent medicine or not; FamHist-family history (number of psychiatric cases in family of the patient).

gradually. On the last day of the washout period and at the end of subsequent treatment weeks, the protocol required various assessments to be performed, of which the HAMD score is used as the response here.

The clinical and demographic background information collected included sex, age, occupation level, illness duration, diagnosis, family history, previous medication, and concurrent medication. Although reasons for dropout are available, these will not be considered here for lack of space.

Full response data are only available for weeks 0, 1, 2, and 4. By this time, 36.4% had dropped out. The results of fitting the dropout models are shown in Table 4. There is no indication of a direct dependence of dropout on treatment. However, it does depend on response, which, in turn, depends on treatment: risk of dropout is greater with higher scores. As in the previous examples, the hazard function has the form of a normal density. Of the covariates, dropout depends on whether concurrent medicine was being taken, the risk being lower if it was, and on family history, being greater with more psychiatric cases in the family. Dropping out also depends on the trial center, with higher rates at centers 2, 4, 5, and 6.

As in the previous example, we must conclude that dropping out is nonignorably nonrandom, but here for quite different reasons. Because of the complex dependence on various covariates, including previous response, it is difficult to plot a typical hazard curve.

5. Discussion

5.1 Two Philosophies of Missing Data

The approach presented here is based on the principle that the dropout process should *always* be modeled, and that it has strategic interest in its own right. On

the other hand, the repeated measurements generally only have meaning conditional on not dropping out. In contrast, the philosophy of those following the Little and Rubin (1) school for missing data, as applied to dropouts, is to model the repeated measurements marginally, as if no one had dropped out. For them, the goal is to determine when one can *avoid* modelling the dropout process. The ignorability criteria in the two cases are very different; those presented here are much stricter and will rarely be fulfilled.

The Little and Rubin (1) distinctions deal essentially with questions of inference about the ''complete'' data distribution, not the appropriateness of models to a given context. Their nonrandom (that is nonindependent) missingness means that this process must be explicitly modelled for any inferences to be possible. With their random (that is independent of unobserved data) missingness, direct likelihood (as used above) and Bayesian inferences are possible while ignoring the missing data process. Only when the data are missing completely at random (that is, independent of both observed and unobserved data) are frequentist inferences possible without this process.

These criteria are widely misunderstood in the literature; frequentist inferences, even based on a likelihood function, and including the usual interpretation of standard errors, are impossible without modeling the dropout process, except in the third case. The probability statements they make rely on the sample space being defined in such a way that it includes anything that might possibly have been observed, thus excluding the second case; strictly speaking, *no* frequentist inferences are possible unless the nonindependent missing data process is modeled, even when there are no missing data among the observations that one had originally planned to record (Lindsey (39, p. 316)). In conclusion, these distinctions provide no guidelines as to what are the appropriate models for the given question under study.

The crucial assumptions in some of these models, those that make the probability of dropout depend on the unobserved future responses, are that the joint distribution of observed responses does not depend on whether or not a subject is about to drop out and that this same joint distribution would continue to hold if further observations had been possible (Diggle and Kenward (9)). These are almost always unrealistic for several reasons. One may expect at least some aspects of the previous history of those about to drop out to differ from those remaining; otherwise, the dropout process is not predictable. Then, once an individual has dropped out, a change of state has occurred so that the distribution of unrecorded longitudinal responses, where considering them makes sense, should be different. Both of these points have been used here in constructing the models; the latter will be discussed further below.

If the first assumption takes the extreme form that the distribution of longitudinal responses has already changed well before dropout, then one can argue for the complete factorization as a pattern mixture model, $f(\mathbf{y}|\mathbf{d})f(\mathbf{d})$, mentioned

above, but from what time point on? In addition, this poses the problem of creating a stratification among the present distributions of longitudinal responses based on a criterion that can only be observed in the future. As well, those in the literature who have advocated this approach integrate over the hypothetical unrecorded future observations after dropout to obtain the marginal distribution, as if dropping out had not occurred; see Little (12) or Hogan and Laird (25).

5.2 Dropouts in Clinical Trials

The common dictum that one should avoid missing values wherever possible by taking energetic steps to retain subjects in a study is not as obvious as it is often thought to be. The advice to have *no* dropouts is not a good guideline if they would be nonrandom in the sense defined here, because it generally imposes unrealistic conditions on the subjects, hence biasing the final conclusions. Either only very untypical subjects, believed not to be prone to dropout, are allowed to enter the trial, or some subjects are forced to remain in the trial against their will and will react differently than the others.

Instead, the dropout process(es), under suitably prespecified conditions, should be accepted as an integral part of the phenomenon under study (although certainly not encouraged!) and modeled as such. Shih and Quan (44) take a similar position in the context of testing the endpoint of a clinical trial; they suggest a joint test for treatment difference in dropping out and in final response conditional on not dropping out.

In any study, the time of dropout should be recorded as accurately as possible, especially if the times between the longitudinal measures are widely spaced. Different reasons for dropout must be recorded and, where relevant, modeled as distinct risk processes. As for any failure process, all covariates that might be useful in predicting dropouts should be recorded, even if they are not directly relevant to the longitudinal measures process itself. Another possibility would be to have stricter entrance requirements at the beginning of a study so that individuals likely to drop out do not participate in the first place. In the context of the modeling strategies developed here, this provides no advantage and has at least two disadvantages. Less information is collected, especially about the dropout process, and the conclusions are less widely generalizable.

When using the models presented here, great care must be taken in concluding that the dropout process is random, in the sense used in this paper. If the risk of dropout does not appear to depend on time, it may simply be that the appropriate function of time has not been found, although the semiparametric Cox model may cover this possibility. In the same way, the conclusion that the process is ignorably nonrandom must be made cautiously because the appropriate dependence on the observed longitudinal measurements responses and covariates may not have ben discovered.

5.3 Extensions

Several extensions to this dropout model are possible and may be useful in certain circumstances. However, except for the first mentioned below, they usually require a relatively large sample size in order to be estimable.

1. If software for nonlinear Poisson regression is available, any risk function can be fitted, not just those in the proportional hazards family.
2. If there are several reasons for dropping out, competing risks procedures may be appropriate, although with the attendant difficulties that they entail.
3. The above models include ''defective'' survival curves that allow for a (latent) subgroup of the population (among those individuals with complete observations) that is thought not to be susceptible to dropping out. However, in certain situations, a finite mixture model may be necessary in order to model such a subgroup explicitly. Taris (28) uses a Markov chain mixture model in such a context.
4. If the longitudinal measures continue to be recorded after change to the ''dropout'' state, for example when a treatment is abandoned but followup continues (Hogan and Laird (45)), the likelihood function (1) for one individual becomes

$$\mathbf{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}; \mathbf{y}, \mathbf{d}) = \prod_{t=1}^{n} f(y_t|\mathcal{F}_{t-}, d_{t-}; \boldsymbol{\alpha})\{[\lambda(t_+|\mathcal{F}_t; \beta)\ \Delta_t]^{d_{t+}} \exp[-\lambda(t_+|\mathcal{F}_t; \beta)\Delta_t]\}^{\delta}$$

   where δ is an indicator variable, one before the change and zero after; and the state indicator, d_{t+}, retains the value one after the ''dropout'' time. The model for the dropout process is identical to that given above. However, the form of $f(y_t|\mathcal{F}_{t-}, d_{t-}; \boldsymbol{\alpha})$ may differ before and after dropout.
5. This can be further generalized to event histories (Lindsey (2, Ch. 10)) with an alternation between two states or the possibility of several states so that d_{t+} can take on several values. Different risk models will generally be required for each state and again the form of $f(y_t|\mathcal{F}_{t-}, d_{t-}; \alpha)$ may be substantially different, depending on the state, that is on the value of d_{t-}. One application would be to noncompliance in clinical trials.

In the light of these extensions, many of the procedures discussed here may sometimes be directly applicable to certain types of missing responses within series of longitudinal measures. Thus, the typology of nonrandomness might still

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

apply so that survival methods could be used to model the missingness process, but now as a series of repeated events. If missing values are not isolated but occur in sequences together, the alternation of two risk processes, stopping and beginning again, could be modeled, although only the former could be made to depend on the longitudinal measures responses.

5.4 Drawing Conclusions

Statisticians often look for too simple an answer to the questions being studied. When there are few dropouts in a longitudinal study, they will contain little information and conclusions will depend primarily on the repeated measurements model. However, when many dropouts occur, the model for them may contribute the essential information, dominating the conclusions. Thus, both must be considered.

For example, in a clinical trial, one unique endpoint for treatment difference is commonly examined. However, in the context of longitudinal measurements, things are usually more complex. Some patients may withdraw because they are cured and others because they suffer side effects. These must be distinguished in any dropout model. (For an example, see Lindsey, 1999, pp. 387–389.) Then, there should be three types of conclusions: treatment differences for the probability of cure and for the side effects, as well as differences among the profiles of the repeated responses of those not yet dropping out. Note that all can be causal because they directly depend on randomization. Then, the repeated measures response profiles might show no differences although one treatment is clearly superior because of more cures or fewer side effects. On the other hand, if one treatment has a superior profile but inferior record of cures or side effects, an example of patient-treatment interaction, the relative importance of these would have to be interpreted, and weighed, according to nonstatistical medical criteria.

Acknowledgments

Philippe Lambert provided helpful comments and suggestions on an earlier version. J. Essers kindly provided the data from the antidepressant study.

References

1. Little RJA, Rubin D: Statistical Analysis with Missing Data, John Wiley & Sons, New York, 1987.
2. Lindsey JK: Models for Repeated Measurements, 2nd Ed., Oxford University Press, Oxford, 1999.

3. Diggle PJ: ''Testing for random dropouts in repeated measurement data,'' *Biometrics*, 45, 1255–1258, 1989.
4. Ridout MS: ''Testing for random dropouts in repeated measurement data,'' *Biometrics*, 47, 1617–1621, 1991.
5. Gould AL: ''A new approach to the analysis of clinical drug trials with withdrawals,'' *Biometrics*, 36, 721–727, 1980; 38, 276–278, 1980.
6. Murray GD, Findlay JG: ''Correcting for the bias caused by drop-outs in hypertension trials,'' *Statist. in Med*., 7, 941–946, 1988; 8, 1302–1304, 1988.
7. Heyting A, Tolboom JTBM, Essers JGA: ''Statistical handling of drop-outs in longitudinal clinical trials,'' *Statist. in Med*., 11, 2043–2061, 1992.
8. Kenward MG, Lesaffre E, Molenberghs G: ''An application of maximum likelihood and generalized estimating equations to the analysis of ordinal data from a longitudinal study with cases missing at random,'' *Biometrics*, 50, 945–953, 1994.
9. Diggle PJ, Kenward MG: ''Informative drop-out in longitudinal data analysis,'' *Appl. Statist*., 43, 49–93, 1994.
10. Lesaffre E, Molenberghs G, Dewulf L: ''Effect of dropouts in a longitudinal study: an application of a repeated ordinal model,'' *Statist. in Med*., 15, 1123–1141, 1996.
11. Molenberghs G, Kenward MG, Lesaffre E: ''The analysis of longitudinal ordinal data with nonrandom drop-out,'' Biometrika, 84, 33–44, 1997.
12. Little RJA: ''Modeling the drop-out mechanism in repeated-measures studies,'' *J. Am. Statist. Assoc*., 90, 1112–1121.
13. Woodbury MA, Manton KG, Stallard E: ''Longitudinal analysis of the dynamics and risk of coronary heart disease in the Framingham study,'' *Biometrics*, 35, 575–585, 1979.
14. Manton KG, Woodbury MA, Stallard E: ''Models of the interaction of mortality and the evolution of risk factor distribution: a general stochastic process formulation,'' *Statist. in Med*., 7, 239–256, 1988.
15. Wu MC, Carroll RJ: '' Estimation and comparison of changes in the presence of informative right censoring by modelling the censoring process.'' *Biometrics*, 44, 175–188, 1988.
16. Schluchter MD: ''Methods for the analysis of informatively censored longitudinal data,'' *Statist. in Med*., 11, 1861–1870, 1992.
17. De Gruttola V, Tu XM: ''Modelling progression of CD4-lymphocyte count and its relationship to survival time,'' *Biometrics*, 50, 1003–1014, 1994.
18. Tsiatis AA, De Gruttola V, Wulfsohn MS: ''Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS,'' *J. Am. Statist. Assoc*., 90, 27–37, 1995.
19. Martinussen T, Keiding N: ''The Manton-Woodbury model for longitudinal data with dropouts,'' *Statist. in Med*., 16, 273–283, 1997.
20. Wu MC, Bailey KR: ''Analyzing changes in the presence of informative right censoring caused by death and withdrawal,'' *Statis. in Med*., 7, 337–346, 1988.
21. Wu MC, Bailey KR: ''Estimation and comparison of changes in the presence of informative right censoring: conditional linear model.'' *Biometrics*, 45, 939–955, 1989; correction 46, 88, 1989.
22. Mori M, Woodworth GG, Woolson RF: ''Application of empirical Bayes inference to estimation of rate of change in the presence of informative right censoring,'' *Statist. in Med*., 11, 621–631, 1992.
23. Pawitan Y, Self S: '' Modeling disease marker processes in AIDS,'' *J. Am. Statist. Assoc*., 88, 719–726, 1993.
24. Hogan JW, Laird NM: ''Mixture models for the joint distribution of repeated measures and event times,'' *Statist. in Med*., 16, 239–257, 1997.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

25. Hogan JW, Laird NM: ''Model-based approaches to analyzing incomplete longitudinal and failure time data,'' *Statist. in Med*., 16, 259–272, 1997.
26. Mori M, Woolson RF, Woodworth GG: ''Slope estimation in the presence of informative right censoring: modeling the number of observations as a geometric random variable,'' *Biometrics*, 50, 39–50, 1994.
27. Fisher RA: ''Presidential Address, First Indian Statistical Conference 1938,'' *Sankhyā*, 4, 14–17, 1938.
28. Taris TW: ''Modeling nonresponse in multiwave panel studies using discrete-time Markov models,'' *Quality and Quantity*, 30, 189–203, 1996.
29. Taris TW: ''On selectivity of nonresponse in discrete-time multi-wave panel studies,'' *Quality and Quantity*, 31, 79–94, 1997.
30. Rubin DB: ''Inference and missing data,'' *Biometrika*, 63, 581–592, 1976.
31. Kalbfleisch JD, Prentice RL: *The Statistical Analysis of Failure Time Data*, John Wiley & Sons, New York, 1980.
32. Kenward MG, Molenberghs G: ''Likelihood based frequentist inference when data are missing at random,'' *Statist. Sci*., 13, 236–247, 1998.
33. Kenward MG, Molenberghs G: ''Parametric models for incomplete continuous and categorical longitudinal data,'' *Statist. Methods in Med. Res*., 8, 51–83, 1999.
34. Lindsey JK: ''Fitting parametric counting processes by using long linear models,'' *Appl. Statist*., 44, 201–212, 1995.
35. Cox DR: ''Regression models and life-tables,'' *J. Roy. Statist. Soc*., B34, 187–220, 1972.
36. Lindsey JK: ''Parametric multiplicative intensities models fitted to bus motor failure data,'' *Appl. Statist*., 46, 245–252, 1997.
37. Yashin AI, Manton KG: ''Effects of unobserved and partially observed covariate processes on system failure: a review of models and estimation strategies,'' *Statist. Sci*., 12, 20–34, 1997.
38. Lindsey JK: ''Construction and comparison of statistical models,'' *J. Roy. Statist. Soc*., B36, 418–425, 1974.
39. Lindsey JK: *Parametric Statistical Inference*, Oxford University Press, Oxford, 1996.
40. Akaike H: ''Information theory and an extension of the maximum likelihood principle,'' In: Second International Symposium on Inference Theory, Petrov BN, Csàki F, eds., Akademiai Kiadó, Budapest, pp. 267–281, 1973.
41. Lindsey JK, Jones B: ''Choosing among generalized linear models applied to medical data,'' *Statist. in Med*., 17, 59–68, 1998.
42. Verbeke G, Lesaffre E: ''The effect of drop-out on the efficiency of longitudinal experiments,'' *Appl. Statist*., 48, 363–375, 1999.
43. Thall PF, Lachin JM: ''Analysis of recurrent events: nonparametric methods for random-interval count data,'' *J. Am. Statist. Assoc*., 83, 339–347, 1988.
44. Shih WJ, Quan H: ''Testing for treatment differences with dropouts present in clinical trials—a composite approach,'' *Statist. in Med*., 16, 1225–1239, 1997.
45. Hogan JW, Laird NM: ''Intention-to-treat analyses for incomplete repeated measures data,'' *Biometrics*, 52, 1002–1017, 1996.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016