Prospective Analysis of Logistic Case-Control Studies

R. J. Carroll, Suojin Wang, and C. Y. Wang

In a classical case-control study, Prentice and Pyke proposed to ignore the study design and instead base estimation and inference on a random sampling (i.e., prospective) formulation. We generalize this prospective formulation of case-control studies to include multiplicative models, stratification, missing data, measurement error, robustness, and other examples. The resulting estimates, which ignore the case-control study aspect and instead are based on a random-sampling formulation, are typically consistent for nonintercept parameters and are asymptotically normally distributed. We derive the resulting asymptotic covariance matrix of the parameter estimates. The covariance matrix obtained by ignoring the case-control sampling scheme and using prospective formulas instead is shown to be at worst asymptotically conservative and asymptotically correct in a variety of problems; a simple sufficient condition guaranteeing the latter is obtained.

KEY WORDS: Asymptotics; Corrections for attenuation; Differential measurement error; Estimating equations; Measurement error; Missing data; Robust estimates.

1. INTRODUCTION

In a classical prospective logistic regression study, a random sample from a source population is taken and the status of a binary outcome $D$ is ascertainment, along with the values of covariates $(Z, X)$, these being related via the logistic regression model

$$Pr(D = 1 | Z, X) = H(\theta_0^* + \theta_1^* Z + \theta_2^* X), \quad (1)$$

where $H(\cdot)$ is the logistic distribution function. The classical case-control study (choice-based sample in econometrics) begins with the model (1), but instead uses retrospective sampling. Specifically, one first obtains a set of cases ($D = 1$) and controls ($D = 0$), and then samples from within the cases and controls to observe the covariates. The analysis of case-control studies of this type was described by Prentice and Pyke (1979), who showed that if one ignored the case-control sampling scheme and analyzed the data as if it came from a prospective sampling scheme, then the resulting estimates of $\theta_1$ and $\theta_2$ are consistent and the usual standard errors are asymptotically correct.

For prospective logistic regression studies, many other types of analyses and sampling schemes are possible. Here are a few examples:

- In problems with partially missing data, one can use likelihood techniques (Little and Rubin 1987) or unbiased estimating equations due to Robins, Rotnitzky, and Zhao (1994).

Although the prospective analyses of these prospective techniques have been worked out, there is to date no corresponding general theory for whether they even lead to consistent estimates when applied to case-control studies and, if they do, whether these prospectively calculated standard errors are asymptotically correct in case-control studies. Our aim is to provide one version of such a theory, and in particular to answer the question: When can prospective analyses be used in case-control studies without having to adjust for the retrospective sampling structure?

We will show that, in general, using prospectively derived standard errors is at worst asymptotically conservative; that is, the standard errors are at worst too large. In addition, we derive a simple sufficient condition guaranteeing that prospective standard errors are asymptotically correct.

In the Appendix we sketch an informal argument derived from a semiparametric perspective that suggests that prospectively computed standard errors are retrospectively correct whenever the distribution of $(Z, X)$ is left unrestricted. Much of this article is a formalization of this argument, along with consideration of cases that are not so easily categorized. The key feature of our analysis is that we start with a general class of unbiased estimating equations, instead of working with specific examples. The results allow for general patterns of missing data as well as for stratified studies. The asymptotic distribution theory is almost trivial to derive in this general framework, thus facilitating the identification of a simple sufficient condition for checking whether prospectively derived standard errors are asymptotically correct. Our results apply not only to the linear logistic model (1), but also to the multiplicative model of Weinberg and Wacholder (1993).

Here is an outline of the article. Section 2 reviews the known results on estimating equations, estimating functions, and sandwich covariance estimation in prospective studies. Using this background, we provide a simple argument showed...
ing why prospective standard errors are at worst asymptotically conservative when applied to case-control studies. Section 3 defines the general estimating equation framework allowing for missing and mismeasured data. Section 4 states the two main results.

The rest of the article considers important special cases, the results for which are new with one exception. Section 5 considers studies with no missing or mismeasured data. In work generalizing that of Weinberg and Wacholder (1993) for multiplicative models and Wang and Carroll (1993, 1995) for robust logistic estimation, we show that essentially any prospectively motivated estimator can be used retrospectively, with asymptotically correct standard errors.

Further sections deal with problems of missing and mismeasured data. Section 6 applies the general theory to a modification of the unbiased estimating equations proposed by Robins et al. (1994) for case-control studies with mismeasured data when there is a validation subsample, allowing for differential measurement error (formally defined in Sec. 6). Section 7 considers measurement error models with nondifferential measurement error in which a validation study can be done, using the simple prospective likelihood methods due to Satten and Kupper (1993). Section 8 discusses measurement error models when validation is not possible, and uses prospective correction-for-attenuation methods. In all three cases, prospective standard errors are asymptotically correct retrospectively.

Section 9 investigates the theory for the partial questionnaire design of Wacholder, Carroll, Pee, and Gail (1994), which has a nonmonotone pattern of missingness; it is shown that, in principle at least, prospective standard errors are asymptotically conservative. The results in Sections 5–9 are new. Section 10 studies the two-stage studies of Breslow and Cain (1988).

2. ESTIMATING EQUATIONS, SANDWICH ESTIMATORS, AND THE CLASSICAL MODEL

One of our main results is that prospectively derived standard errors are at worst asymptotically conservative. Justification for this result is easiest to understand in the classical simple logistic model, \( \text{pr}(D = 1 | X) = H(\theta_0 + \theta_1 X) \). The argument uses nothing more than standard estimating equation theory; we will outline this theory in the nomenclature as we go along. Extensions to complex problems require little more than change in notation.

2.1 Prospective Sampling

We first consider prospective sampling, and write \( \Theta^* = (\theta_0^*, \theta_1^*)' \). The prospective ordinary logistic regression estimate is the solution to the equation

\[
0 = \sum_{i=1}^{n} \left( \frac{1}{X_i'} \right) \left[ D_i - H(\theta_0^* + \theta_1^* X_i) \right].
\]

The entire term on the right side of (2) is called an estimating equation. The arguments \( \psi(D_i, X_i, \Theta^*) \) are called estimating functions. The prospective estimator is denoted by \( \hat{\Theta}^* \).

Prospective theory requires that the estimating equation be unbiased; that is, it has mean zero when evaluated at the parameters, so that

\[
0 = E\left\{ \sum_{i=1}^{n} \psi(D_i, X_i, \Theta^*) \right\}. \tag{3}
\]

For logistic regression, even more is true. The estimating functions are themselves unbiased, having mean zero at the parameters:

\[
0 = E\{ \psi(D_i, X_i, \Theta^*) \} \quad \text{for} \quad i = 1, \ldots, n. \tag{4}
\]

However, only Equation (3) is required for now.

By use of Taylor series, it is known that \( \hat{\Theta}^* \) is asymptotically normally distributed (under regularity conditions), and we write the distribution as

\[
\sqrt{n}(\hat{\Theta}^* - \Theta^*) \approx \text{Normal} \{0, B^{-1}(\Theta^*)A(\Theta^*)B^{-1}(\Theta^*)\}, \tag{5}
\]

where

\[
B(\Theta^*) = n^{-1} \sum_{i=1}^{n} E\left\{ \frac{\partial}{\partial \Theta^*} \psi(D_i, X_i, \Theta^*) \right\} \tag{6}
\]

and

\[
A(\Theta^*) = n^{-1} \text{cov} \left\{ \sum_{i=1}^{n} \psi(D_i, X_i, \Theta^*) \right\}. \tag{7}
\]

Formula (5) is often called the sandwich formula, because \( A(\Theta^*) \) is sandwiched between inverses of \( B(\Theta^*) \).

At this point, we may now use the fact that the estimating functions are unbiased—that is, use (4)—to conclude that if we define

\[
n^{-1} \sum_{i=1}^{n} E\{ \psi(D_i, X_i, \Theta^*) \psi(D_i, X_i, \Theta^*)' \} = C(\Theta^*), \tag{8}
\]

then \( A(\Theta^*) = C(\Theta^*) \) and \( \hat{\Theta}^* \) is asymptotically normally distributed with mean \( \Theta^* \) and covariance matrix

\[
n^{-1} B^{-1}(\Theta^*) C(\Theta^*) B^{-1}(\Theta^*). \]

Of course, in ordinary logistic regression we know that \( C(\Theta^*) \) and \( B(\Theta^*) \) are equal and can be consistently estimated by the usual information formula. In general, though, a consistent nonparametric estimate of these terms can be based on the method of moments; that is, in (6) and (8) remove the expectations and replace \( \Theta^* \) by \( \hat{\Theta}^* \). The resulting covariance matrix estimator is sometimes called the robust sandwich formula, where in a misnomer the term "robust" is used as a replacement for "model-free" (Drum and McCullagh 1993). For example, the resulting model-free estimate of \( B(\Theta^*) \) is just

\[
\hat{B}(\Theta^*) = n^{-1} \sum_{i=1}^{n} \frac{\partial}{\partial \Theta^*} \psi(D_i, X_i, \hat{\Theta}^*). \]

2.2 Retrospective Sampling

We now turn to retrospective sampling. The key point to notice in the preceding argument is that we used unbiasedness of the estimating functions only in showing that (7) equals (8).

In retrospective sampling, define \( \theta_0 = \theta_0^* + \log(n_1/n_0) \)}
\[- \log \{ \text{pr}(D = 1) / \text{pr}(D = 0) \}, \text{ where } \text{pr}(D = 1) \text{ is the unknown prospective rate. Prentice and Pyke (1979) showed that if } \Theta = (\theta_0, \theta_1)^T \text{ and we replace } \Theta^* \text{ by } \Theta, \text{ then the estimating function (2) is still unbiased; that is, (3) holds. But the estimating functions are not unbiased, so that (4) fails, and hence it is not true that } A(\Theta) = C(\Theta).\]

Asymptotically, the distribution (5) still remains the same, but with the prospective parameter \( \Theta^* \) and prospective estimator \( \hat{\Theta}^* \) replaced by the retrospective parameter \( \Theta \) and retrospective estimator \( \hat{\Theta} \), which of course is the solution to (2) under retrospective sampling.

Because the estimating equation is unbiased, we can re-write (7) as follows:

\[
A(\Theta) = n^{-1} \text{cov} \left[ \sum_{i=1}^n \psi(D_i, X_i, \Theta) - E \left[ \sum_{i=1}^n \psi(D_i, X_i, \Theta) \right] \right] \\
= n^{-1} \sum_{i=1}^n \text{cov} \{ \psi(D_i, X_i, \Theta) - E \{ \psi(D_i, X_i, \Theta) \} \} \\
= n^{-1} \sum_{i=1}^n E \{ \psi(D_i, X_i, \Theta) \} \psi^*(D_i, X_i, \Theta) \\
- n^{-1} \sum_{i=1}^n E \{ \psi(D_i, X_i, \Theta) \} E \{ \psi^*(D_i, X_i, \Theta) \} \\
= C(\Theta) - D(\Theta).
\]

The main conclusion now follows, in a series of steps:

- Prospectively, the asymptotic covariance matrix is \( n^{-1} B^{-1}(\Theta^*) C(\Theta^*) B^{-T}(\Theta^*) \).
- Applying prospective formula directly to a retrospective study is equivalent to biasing estimation as if the correct covariance matrix were \( n^{-1} B^{-1}(\Theta) C(\Theta) B^{-T}(\Theta) \).
- But the proper covariance is \( n^{-1} B^{-1}(\Theta) \{ C(\Theta) - D(\Theta) \} B^{-T}(\Theta) \).
- Because \( D(\Theta) \) is positive semidefinite, prospective covariance formulas are at worst conservative.

2.3 Further Steps

The reasoning given previously is perfectly sound, but we have skipped over a few steps. For example, we have simply assumed that the actual covariance estimators derived from a prospective analysis estimate the corresponding quantities retrospectively, which is true but needs to be justified.

Our analysis shows that prospective covariance formulas are at worst conservative, but no insight is given as to when these formulas are asymptotically correct. Our second main contribution is to derive a simple sufficient condition for this asymptotic correctness. The condition is routine to check in the examples described in this article, as well as other examples that we have not included. Deriving the sufficient condition requires a more detailed examination of \( D(\Theta) \).

This task is relegated to the general theory.

3. PROSPECTIVE FORMULATION

3.1 Likelihood for Complete Data

The following conventions are used throughout. Disease status is denoted by \( D \), observable covariates by \( Z \), and covariates that may be partially missing by \( X \). Anticipating the possibility that the study may be stratified, we use the stratification variable \( S \), taking on the values \( 1, \ldots, s \). When considering measurement error problems, instead of observing \( X \), a proxy \( W \) is typically observed for all study participants; for example, blood pressure measured at a single time point as a proxy for long-term blood pressure measurement.

The vector of parameters of major interest is denoted by \( \theta_1 \), for example, in (1), \( \theta_1 = (\theta_1', \theta_2')' \).

If there were no missing data, then we assumed a sampling mechanism of a classical case-control study within each stratum \( S = s \), with \( n_{1s} \) cases, \( n_{0s} \) controls and \( n_s = n_{1s} + n_{0s} \) observations. The total sample size is \( n = \sum n_s \). We assume that the terms \( n_s/n \) converge to positive constants, so that our work does not apply to matched case-control studies.

In those cases where a proxy \( W \) exists, it is sometimes useful to allow for an error model for it. Thus we write the likelihood of \( W \) given \( (D, Z, X) \) and stratum \( S = s \) as \( f(w|z, x, d, s, \theta_2) \), where \( \theta_2 \) is an unknown parameter. We will assume that the prospective model is of the form

\[
\text{pr}(D = 1|Z, X, S = s) = H(\theta_0^* + R_s(\theta_1, \theta_2, Z, X)),
\]

where \( R_s(\theta_1, \theta_2, Z, X) \) is an arbitrary function. Although the vector \( \theta_2 \) is in both the conditional likelihood for \( W \) and in (9), this is simply a convention; not all components of \( \theta \) must appear in both likelihoods. Model (9) includes the linear logistic model (1) and the multiplicative model of Weinberg and Wacholder (1993) as special cases.

From the usual odds ratio formulation of Prentice and Pyke (1979), the retrospective likelihood that \( (Z, W, X) = (z, w, x) \) when \( (D, S) = (d, s) \) is

\[
(n_s/n_{0s}) q_s(z, x) H_s^*(\cdot) \{ 1 - H_s(\cdot) \} ^{1-d} f(w|z, x, d, s, \theta_2),
\]

where

\[
H_s(\cdot) = H(\theta_0^* + R_s(\theta_1, \theta_2, Z, X)).
\]

In (10), \( q_s(\cdot) \) is the marginal density of \((Z, X) \) in stratum \( S = s \) induced by the case-control sampling scheme and \( \theta_0^* = \theta_0^* + \log(n_{0s}/n_{0s}) - \log \{ \text{pr}(D = 1|S = s)/\text{pr}(D = 0|S = s) \} \), where \( \text{pr}(D = 1|S = s) \) is the prospective rate in stratum \( s \). We write \( \Theta = (\theta_{c1}, \ldots, \theta_{c2}, \theta_1', \theta_2')' \), the retrospective parameter, and \( \Theta^* = (\theta_{01}, \ldots, \theta_{02}, \theta_1', \theta_2')' \), the prospective parameter.

3.2 Missing Data

The theory allows for the possibility that different components of \( X \) are missing in different subsets of the data. If there are \( J \) such possible patterns of missingness \( \Delta = (\delta_1, \ldots, \delta_J) \) is a vector with a single nonzero component indicating which pattern is applicable. The only assumption is that the data are missing at random, and hence the missing data indicators and \( X \) are conditionally independent given \((Z, W, S, D) \), with selection probabilities \( \pi_j(Z, W, S, D) = \text{pr}(\delta_j = 1|Z, W, S, D, X) \).

For example, suppose that \( X \) has two components, \( X_{(1)} \) and \( X_{(2)} \). There are four possible patterns of missingness here: (1) both components missing; (2) only \( X_{(1)} \) missing;
(3) only $x_{(j)}$ missing; and (4) neither component missing. In this case, $\delta_1 = 1$ means that both components of $X$ are missing, $\delta_2 = 1$ means that only $x_{(1)}$ is missing, and so on. Table 1 summarizes the notation.

3.3 Prospective Estimating Equations

With the exception of the leading term, (10) is of the same general form as a prospective likelihood with stratum-specific intercepts. Thus a natural approach to estimation is to use prospective estimating equations. Let $\delta_{ij}$ denote the value of $\delta_j$ for the $i$th individual in the $j$th stratum. The prospective estimating function defined for the $j$th pattern of missingness and the $s$th stratum is $\Psi_{js}(D, Z, X, W, s, \Theta)$, and the estimators are defined as solutions to

$$0 = n^{-1} \sum_{s=1}^{g} \sum_{i=1}^{n_s} \sum_{j=1}^{f} \delta_{ij} \Psi_{ij}(D_{it}, Z_{it}, X_{it}, W_{it}, s_{i}, \Theta)$$

$$= n^{-1} \sum_{s=1}^{g} \sum_{i=1}^{n_s} \log L_{n}(\Theta) = T_{n}(\Theta).$$

(11)

In effect, we are suggesting that one ignore the case-control study design and proceed as if the data arose from a prospective sample.

4. ASYMPTOTIC THEORY

4.1 Main Results

Readers who are interested mainly in the applications may skip this section without any loss.

In our analysis we make two basic assumptions. First, we assume that given $(D, S = s)$, the vectors $(Z_{it}, X_{it}, W_{it}, \Delta_{it})$ are independent and identically distributed as $i$ varies. The individual components of these vectors are, of course, dependent. The assumption of independent and identically distributed data is only for simplicity in this analysis and is not always necessary, as we show in Section 10.

The second assumption is that Equation (11) is retrospectively unbiased, so that

$$0 = \sum_{s=1}^{g} \sum_{i=1}^{n_s} E \{ \log L_{n}(\Theta) | D_{it}, s \}. $$

(12)

Assumption (12) is satisfied in all the cases that we have examined. As described in more detail in the Appendix, Section A.2, this appears to be a general phenomenon and not simply a matter of convenient example selection on our part.

To state the main result, we set the following definitions. Define $l_{\star}(\cdot, \Theta) = H_{\star}^{-1}(\cdot) | 1 - H_{\star}(\cdot) |^{1/2} f(w, z, x, d, s, \theta_2) q_{\star}(\cdot)$. The notation $d_\theta (\cdot)$ means integration or summation with respect to the arguments of $\mu(\cdot)$. Let $\Psi_{js\theta}$ be the matrix of partial derivatives of $\Psi_{js}$ with respect to $\Theta$.

Also, define

$$T_{\Theta}(\Theta) = \sum_{s=1}^{g} (n_s / n) \sum_{d=0}^{f} \sum_{j=1}^{f} \int \pi_{j}(\cdot) \Psi_{j\theta}(\cdot, \Theta)$$

$$\times l_{\star}(\cdot, \Theta) d_\theta (w, z, x, w),$$

(13)

and

$$\kappa_s = \int \sum_{s=1}^{f} \pi_{j}(\cdot) \Psi_{js}(\cdot, \Theta) l_{\star}(\cdot, \Theta) d_\theta (z, x, w).$$


<table>
<thead>
<tr>
<th>Variable</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$</td>
<td>Response</td>
</tr>
<tr>
<td>$Z$</td>
<td>Fully observed covariates</td>
</tr>
<tr>
<td>$X$</td>
<td>Missing or mismeasured covariates</td>
</tr>
<tr>
<td>$W$</td>
<td>Proxy for $X$ in measurement error problems</td>
</tr>
<tr>
<td>$S$</td>
<td>Stratum indicator variable</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>Indicator that $X$ is missing with pattern number $j$</td>
</tr>
<tr>
<td>$\pi_{j}(z, w, s, \Theta)$</td>
<td>Probability of missing data pattern $j$</td>
</tr>
<tr>
<td>$\Theta$</td>
<td>Retrospective parameter, including stratum intercepts</td>
</tr>
<tr>
<td>$\Theta^*$</td>
<td>Prospective parameter, including stratum intercepts</td>
</tr>
<tr>
<td>$\kappa_s$</td>
<td>Non-intercept parameter in the prospective logistic model</td>
</tr>
<tr>
<td>$\kappa_{1\Theta}$</td>
<td>Error model parameter for the distribution of $W$</td>
</tr>
</tbody>
</table>

Theorem. Let $\hat{\Theta}$ be the solution to (11) under retrospective sampling, and let $\hat{\Theta}^*$ be the solution to (11) under prospective sampling. Under appropriate regularity conditions, retrospectively, $n^{1/2}(\hat{\Theta} - \Theta)$ is asymptotically normally distributed with mean zero and covariance matrix

$$\{ T_{\Theta}(\Theta) \}^{-1} \left[ C(\Theta) - \sum_{s=1}^{g} \sum_{d=0}^{f} \left\{ n_s / \{ n_s \} \right\} \kappa_s \kappa_{1\Theta} \right] \times \{ T_{\Theta}(\Theta) \}^{-1}. $$

(15)

Prospectively, define $l_{\star}(\cdot, \Theta^*) = q_{\star}(z, x) H_{\star}^{1/2}(\cdot) \{ 1 - H_{\star}(\cdot) \}^{1/2} f(\cdot, z, x, d, s, \theta_2) q_{\star}(\cdot)$.

The notation $d_\theta (\cdot)$ means integration or summation with respect to the arguments of $\mu(\cdot)$. Let $\Psi_{js\theta}$ be the matrix of partial derivatives of $\Psi_{js}$ with respect to $\Theta$.

Also, define

$$T_{\Theta}(\Theta) = \sum_{s=1}^{g} (n_s / n) \sum_{d=0}^{f} \sum_{j=1}^{f} \int \pi_{j}(\cdot) \Psi_{j\theta}(\cdot, \Theta)$$

$$\times l_{\star}(\cdot, \Theta) d_\theta (w, z, x, w),$$

(13)

and

$$C(\Theta) = \sum_{s=1}^{g} (n_s / n) \sum_{d=0}^{f} \sum_{j=1}^{f} \int \pi_{j}(\cdot) \Psi_{j\theta}(\cdot, \Theta) \Psi_{j\theta}(\cdot, \Theta)$$

$$\times l_{\star}(\cdot, \Theta) d_\theta (w, z, x, w),$$

(14)

The proof is sketched in the Appendix.

4.2 When are Prospective Standard Errors Asymptotically Correct?

For the moment, assume that prospectively derived covariance matrix estimates are consistent estimates of the quantity

$$\{ T_{\Theta}(\Theta) \}^{-1} C(\Theta) \{ T_{\Theta}(\Theta) \}^{-1}. $$

(17)

If this is true, then the foregoing theorem states that the prospective covariance matrix estimates are at worst conservative.

Here we state a simple sufficient condition that guarantees that prospectively derived standard errors are asymptotically correct. For most cases, $\kappa_{0\Theta} = -\kappa_{1\Theta}$ for each stratum, and we assume this here. This leads to the following simple result.

Corollary. Suppose that $\kappa_{0\Theta} = -\kappa_{1\Theta}$ and that $\kappa_{1\Theta}$ is proportional to the $s$th column of $T_{\Theta}(\Theta)$ for $s = 1, \ldots, g$. Then prospectively derived covariance formulas for $(\hat{\Theta}, \hat{\Theta}_{1\Theta})$ are
asymptotically correct. More generally, the result holds if the rows of $T_{\theta}(\theta)\kappa_t$ corresponding to $\theta$, all equal zero.

We show later that many examples satisfy the conditions of this corollary.

The reason that prospectively derived covariance matrix estimators actually estimate (17) is that they are in all circumstances derived from sums of functions of $\hat{\theta}$ and the individual observations. For example, consider the model-free sandwich estimator from prospective formulas (Sec. 2), namely

$$\{T_{\theta}(\hat{\theta})\}^{-1} n^{-1} \sum_{s=1}^{S_t} \sum_{i=1}^{n_s} \frac{\partial}{\partial \theta} \ell_{si}(\hat{\theta}) \{T_{\theta}(\hat{\theta})\}^{-1},$$

where

$$T_{\theta}(\hat{\theta}) = n^{-1} \sum_{s=1}^{S_t} \sum_{i=1}^{n_s} \frac{\partial}{\partial \theta} \ell_{si}(\theta).$$

Using the retrospective likelihood (10) and the fact that $\hat{\theta}$ is a consistent estimator of $\theta$, it is easily seen that the model-free sandwich estimator consistently estimates (17).

For those cases where for prospective formulas are conservative, there are two ways to construct asymptotically correct covariance estimates. The preferred method is to begin with (15) and estimate $T_{\theta}(\hat{\theta})$ and $C(\theta)$ by prospective formulas; typically, one would not use the "model-free" estimates of these terms. For example, in the classical problem with no missing data, these matrices would be estimated by the observed information. To estimate $n_{dt}$ in (15), use $n_{dt} = n^{-1} \sum_{i=1}^{n_s} I(D_{it} = d) \ell_{si}(\hat{\theta})$, a model-free consistent estimate.

This hybrid approach, where $n_{dt}$ is estimated without a model and $T_{\theta}(\hat{\theta})$ and $C(\theta)$ are typically based on a prospective model, will work for most cases. But it need not yield a positive semidefinite covariance matrix estimate, because of the subtraction in (15). In such cases, a model-free sandwich covariance matrix estimate can be employed, namely $\{T_{\theta}(\hat{\theta})\}^{-1} B_n(\hat{\theta}) \{T_{\theta}(\hat{\theta})\}^{-1}$, where

$$B_n(\hat{\theta}) = n^{-1} \sum_{s=1}^{S_t} \sum_{d=0}^{D} I(D_{is} = d) \times \left[ \ell_{si}(\hat{\theta}) - \hat{m}(d, s, \theta) \right] R \left[ \ell_{si}(\hat{\theta}) - \hat{m}(d, s, \theta) \right],$$

where $\hat{m}(d, s, \theta) = n_{ts}^{-1} \sum_{i=1}^{n_s} I(D_{is} = d) \ell_{si}(\hat{\theta})$ is an estimate of $E[\ell_{si}(\theta) | D_{is} = d]$.

5. CLASSICAL STUDIES

By a classical case-control study, we mean one with no missing data and a single stratum. Dropping the substrates $(j, s)$, which indicate missing data pattern and stratum number, from (9) we have $Pr(D = 1 | X) = H(\theta^n_0 + R(\theta_1, X))$. In this section we show that in classical case-control studies, essentially any reasonable prospectively defined estimating equation yields consistent estimators, and the prospective standard errors are asymptotically correct. The work generalizes that of Weinberg and Wacholder (1993) on multiplicative models and Wang and Carroll (1993, 1995) on robust estimation.

To motivate the class of estimators, first consider simple linear logistic regression with $R(x, \theta_t) = \theta_t x$, and recall from (2) that the estimating function for the maximum likelihood estimator is $\psi(d, x, \theta^*_t) = (1, x)^t \{d - H(\theta^n_0 + \theta_1 x)\}$. By assumption, prospectively

$$E[\psi(D, X, \theta^*_t) | X] = 0,$$

because $Pr(D = 1 | X) = H(\theta^n_0 + \theta_1 X)$. For the prospective maximum likelihood estimator in the general model, the estimating equation for the maximum likelihood estimator is $\psi(d, x, \theta^*_t) = (1, x)^t \{d - H(\theta^n_0 + R(\theta, x))\}$, and (18) still holds. The same condition applies to all the robust estimators discussed by Carroll and Peterson (1993).

The fact then is that most estimators prospectively satisfy (18). We will say that an estimating function $\psi(D, X, \theta^*_t)$ is (prospectively) conditionally unbiased if (18) holds prospectively for all $\theta^*_t$.

In the Appendix, we show the following result.

Lemma. Any conditionally unbiased estimating function leads to a prospectively unbiased estimating equation, and prospectively derived standard errors are asymptotically correct.

The result is anticipated from the Appendix Section A.1, because in this context no restrictions have been made on the marginal distribution of $X$.

5.1 A Simulation

We performed a small simulation in simple linear logistic regression to illustrate the results. There were 75 cases and 75 controls. The predictor $X$ was generated either as a normal random variable with mean zero and variance 1 or as a t-distributed random variable with 3 degrees of freedom. We chose $\theta^n_0 = -4.0$, $\theta_1 = -.4, -.6, -8$. When $X$ is normally distributed, the values of $\theta_1$ were chosen so that the relative risks of moving from the 90th to the 10th percentile of the distribution of $X$ equal 3, 5, and 8. There were 500 simulations for each case.

Two prospectively derived estimators were considered: (1) the ordinary linear logistic estimator, and (2) the robust leverage-downweighting estimators defined by Carroll and Pederson (1993, sec. 4.1). The results are given in Table 2. Note that in all cases, both the ordinary and the robust methods attain very nearly their nominal levels.

| Table 2: Simulation of Ordinary and Robust Logistic Regression |
|----------------|----------------|----------------|----------------|----------------|----------------|
|                |                | 90%            | 95%            | Median         | 90%            | 95%            | Median         |
| Distribution   | $\theta_1$    |                |                |                |                |                |                |
| Normal         | -.4            | .888           | .944           | -335           | .888           | .940           | -434           |
|                | -.6            | .878           | .940           | -610           | .864           | .940           | -612           |
|                | -.8            | .912           | .964           | -816           | .912           | .964           | -806           |
| t(3)           | -.4            | .912           | .964           | -400           | .914           | .964           | -403           |
|                | .6             | .888           | .936           | -618           | .882           | .924           | -619           |
|                | -.8            | .906           | .962           | -812           | .986           | .962           | -811           |

NOTE: In 500 simulations, the coverage rates are given for nominal 90% and 95% intervals. The median of the slope estimates is also listed.
6. MISMEASURED DATA: DIFFERENTIAL ERROR

6.1 Introduction

In most problems with missing data, and less frequently in problems with measurement error, \( X \) is observable in a subset of the study. A wide variety of parametric techniques have been developed for likelihood analysis of missing data, and the corresponding likelihoods for measurement error models are also well known. Recently, however, techniques have been developed which attempt to avoid strong parametric assumptions (see, for example, Carroll and Wand 1991, Pepe and Fleming 1991, and Reilly and Pepe 1995).

We will say that measurement error is **nondifferential** and that \( W \) is a **surrogate** for \( X \) if \( W \) is independent of \( D \) given \((Z, X, S)\). Otherwise, measurement error is **differential**.

Robins et al. (1994) described a general class of prospectively unbiased estimating equations for missing and mis-measured data in a single stratum. We concentrate on this case in linear logistic regression and by modifying their approach slightly allow for differential measurement error. As a matter of interpretation, we take the view that interest lies in the effects of \( X \) on disease in the presence of the covariates \( Z \) measured without error, and not otherwise in \( W \). Thus the interesting prospective logistic model is \( H(\theta_0^* + \theta_1^* Z + \theta_2^* X) \). Our analysis requires a model for the error distribution of \( W \) given \((Z, X, D)\).

6.2 Estimating Equations and Results

The estimating equations can be described as follows. Let \( \psi(d, x, z, \theta) \) be the usual logistic estimating function \( M(z, x)[d - H(\cdot)] \), where \( M(z, x) = (1, z', x') \). Write the conditional density or mass function for \( W \) as \( f(w|z, x, d, \theta) = f(w|z, x, d, \theta_2) \). Let \( \chi(z, x, w, d, \theta) \) be any unbiased estimating function for \( \theta_2 \).

For any function \( \xi = \xi(z, x, w, d) \), define

\[
\psi_1(\cdot, \theta) = \left[ \begin{array}{c}
\psi(D, Z, X, \theta) - \frac{\phi(Z, X, W, \pi \psi, \theta)}{\phi(Z, X, W, \pi, \theta)} - \frac{\phi(Z, X, W, (1 - \pi) \psi, \theta)}{\phi(Z, X, W, (1 - \pi), \theta)} \\
\chi(Z, X, W, D, \theta) - \frac{\phi(Z, X, W, \pi \chi, \theta)}{\phi(Z, X, W, \pi, \theta)}
\end{array} \right]
\]

and

\[
\psi_2(\cdot, \theta) = \left[ \begin{array}{c}
\phi(D, Z, W) \\
0
\end{array} \right].
\]

Note that because there is a single stratum, we have dropped the index corresponding to stratum assignment. This estimating equation is prospectively unbiased and, as can be verified directly, also retrospectively unbiased (see the Appendix, Sec. A.5). In that section we also show that prospectively derived standard errors are asymptotically correct.

7. LIKELIHOOD AND NONDIFFERENTIAL MEASUREMENT ERROR

Satten and Kupper (1993) considered likelihood analysis for prospective studies with nondifferential measurement error. We study their easily computed "unconditional" method in the context of the logistic model (1), showing that it leads to consistent estimates in the retrospective model and that prospectively derived standard errors are retrospectively asymptotically correct.

Prospectively, Satten and Kupper formulated the problem as follows. For all subjects, \((D, Z, W)\) is observed. But for the \(i\)th individual either \(X_i\) is also observed (\(d_i = 1\)) or \(X_i\) is not observed (\(d_i = 1\)). If \(f_{X|ZW,D}(x|z, w, d = 1, \theta)\) is the density or mass function of \(X\) given \((Z, W, D)\), then the prospective likelihood can be written as

\[
\prod_{i=1}^{n} \left[ f_{X|ZW,D}(x_i|z, w, d = 1, \theta) \right] \left\{ \text{Pr}(D_i = 1|Z_i, W_i) \right\}^{D_i} \times \left\{ 1 - \text{Pr}(D_i = 1|Z_i, W_i) \right\}^{1-D_i}.
\]

(19)

The hard part is to compute each term. Satten and Kupper's approach is to model the distribution of \(X\) given \((Z, W, D = 0)\); that is, among the controls, depending on a parameter \(\theta_2\). Define

\[
R(Z, W, \theta) = \log \left\{ \text{Pr}(D = 1|Z, W) \right\} = H(\theta_0^* + \theta_1^* Z + R(Z, W, \theta)).
\]

Prospectively, they showed that \(\text{Pr}(D = 1|Z, W) = H(\theta_0^* + \theta_1^* Z + R(Z, W, \theta))\), and further that the ratio of the density or mass functions is

\[
\frac{f_{X|ZW,D}(x|z, w, d = 1, \theta)}{f_{X|ZW,D}(x|z, w, d = 0, \theta)} = \exp \left\{ \theta_1^* x - R(z, w, \theta) \right\},
\]

thus writing the conditional density of \(X\) given \((Z, W, D = 1)\) in terms of that of \((Z, W, D = 0)\).

The prospective likelihood (19) is now specified, and the maximum likelihood estimator can be computed. In the Appendix, Section A.6, we show that maximizing this prospective likelihood leads to estimators that are retrospectively consistent and standard errors that are asymptotically correct.

8. MEASUREMENT ERROR AND REPLICATION

8.1 Introduction

The classical formulation of the measurement error problem (Fuller 1987) is one in which the true predictor \(X\) is not
observable, and instead only an unbiased surrogate (defined in Sec. 6) \( W \) for \( X \) may be observed, possibly with replication on a subset of the data. If the variance of the measurement error is known or estimated from external data sources, then the standard linear regression method is the so-called "correction for attenuation." In nonlinear regression models, the same correction for attenuation often works extremely well.

There are a variety of proposals based on the idea of a correction for attenuation (see Carroll and Stefanski 1990; Gleiser 1990; Liu and Liang 1992; Rosner et al. 1989; Rosner, Spiegelman, and Willett 1990; and Schaefer 1993). Carroll and Stefanski (1994) described an instrumental variables method.

These methods differ fundamentally from the moments methods of Section 6 in that they apply in the common case where \( X \) is never observable; for example, blood pressure or diet history.

The application of these ideas to case-control studies with nondifferential measurement error were briefly explored by Rosner et al. (1989), studied by Armstrong, Howe, and Whittemore (1989) and Buonaccorsi (1990) using discriminant analysis techniques and allowing for differential measurement error, and suggested as a general methodology with partially replicated data by Carroll, Gail, and Lubin (1993). Although all of these methods can be analyzed by our general theory, in this section we define and investigate a version of the correction for attenuation methodology based on prospective considerations but appropriate for case-control studies. The asymptotic distribution theory is most naturally studied using two strata.

### 8.2 Estimating Equations and Results

We will assume that \( W \) is a surrogate for \( X \), that is, independent of \( D \) given \((Z, X)\), and that the surrogate can be replicated with independent errors. Let \( W = (W_1, W_2) \), where \( W_j = X + U_j \) and \( U_1, U_2 \) are independent and identically distributed with mean zero and variance \( \sigma^2 \). To keep the analysis simple, we will ignore \( Z \) and study the prospective model \( H(\theta_0 + \theta X) \). There are two strata \((s = 1, 2)\): one in which only \( W_1 \) is observed \((s = 1)\), the other for both \( W_1, W_2 \) are observed \((s = 2)\). Set \( J = 1 \) and \( \pi_1 = 1 \).

A good approximation (Carroll and Stefanski 1990; Gleiser 1990; Rosner et al. 1989) to the probability of response given the observed surrogate is

\[
pr(D = 1 | W_1) = H(\theta_0^* + \theta m_1(W_1))
\]

and

\[
pr(D = 1 | W_1) = H(\theta_0^* + \theta m_2(W))
\]

where \( m_1(W_1) = E(X | W_1) \) and \( m_2(W) = E(X | W) \). The correction-for-attenuation methodology estimates the functions \( m_1, m_2 \) and regresses the response on these estimated functions, with one intercept per stratum.

Of course the regression functions \( m_1, m_2 \) are not estimable, because they depend on the underlying disease rates. But they can be approximated in the common case that the disease is rare, because they are approximately the same in the controls as they are in the source population, and hence

\[
m_1(W_1) \approx E(X | W_1, D = 0) \quad \text{and} \quad m_2(W) \approx E(X | W, D = 0)
\]

approximations that we will henceforth treat as exact. Let \( \mu_w \) be the mean of \( W_1 \) among the controls. Following Carroll and Stefanski (1990) and Gleiser (1990), for \( s = 1, 2 \) estimates of the best linear approximations to these regressions are

\[
g_1(W_1, \hat{\sigma}_w^2, \hat{\sigma}_x^2, \hat{\mu}_w) = \hat{\mu}_w + \frac{\hat{\sigma}_w^2}{\hat{\sigma}_x^2} (W_1 - \hat{\mu}_w)
\]

and

\[
g_2(W, \hat{\sigma}_w^2, \hat{\sigma}_x^2, \hat{\mu}_w)
\]

where \( \hat{\sigma}_x^2 \) is the sample variance of \( W_1 \) among all the controls and \( \hat{\sigma}_w^2 \) is the sample variance of \( (W_1 - W_2)/2 \) among the replicated data.

The algorithm then is as follows. Use the replicated data to construct \( \hat{\sigma}_x^2 \) and use the \( W_1 \)'s from all the control data to construct \( \hat{\mu}_w \) and \( \hat{\sigma}_w^2 \). Then regress \( D \) on the functions \( (g_1, g_2) \) for \( s = 1, 2 \), with stratum specific intercepts.

Subject to the levels of approximation described in the Appendix, Section A.7, we show that prospective covariance formulas may be used for the estimate of \( \theta_1 \). The preceding analysis is readily extended to problems with vector predictors.

### 8.3 A Simulation

We performed a small simulation in simple linear logistic regression to illustrate the results. There were 300 cases and controls. The number of replicated observations in each case and control was 25. Prospectively, the variable \( X \) was generated as a normal random variable with mean zero and variance 1. We chose \( \theta_0 = -4.0, \theta_1 = -4, -6, -8 \). With these values, prospectively the event rates are approximately 3%, the type of "rare disease" situation one might expect. The values of \( \theta_1 \) were chosen so that the relative risks of moving from the 90th to the 10th percentile of the distribution of \( X \) equal 3, 5, and 8.

The measurement error model was \( W = X + U \), where \( U \) is independent of \( D \) and \( X \) and is generated as a normal random variable with mean zero and variance \( \sigma^2 = .25, .5, 1.0 \), representing small, moderate, and large measurement error.

We estimated \( \theta_1 \) using the foregoing algorithm. Standard errors were computed using a prospective method described in the Appendix, Section A.7. The results, given in Table 3, indicate that the prospectively derived confidence intervals achieve very nearly their nominal levels.

### 9. Partial Questionnaires

#### 9.1 Introduction

The partial questionnaire design of Wacholder et al. (1994) is a single-stratum design where the covariate \( Z \) is of primary interest and the components of \( X = (X_1, X_2) \) are partially missing by design. Such designs may be of considerable use when \( X_1 \) and \( X_2 \) are expensive or difficult to assess. The advantage of deliberately making components of \( (X_1, X_2) \) missing is a lesser burden on study subjects, possibly resulting
Table 3. Simulation of Correlation for Attenuation

<table>
<thead>
<tr>
<th>( \sigma^2_0 )</th>
<th>( b_1 )</th>
<th>90%</th>
<th>95%</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>.25</td>
<td>.4</td>
<td>.89</td>
<td>.96</td>
<td>-.41</td>
<td>-.41</td>
</tr>
<tr>
<td></td>
<td>.6</td>
<td>.89</td>
<td>.96</td>
<td>-.61</td>
<td>-.60</td>
</tr>
<tr>
<td></td>
<td>.8</td>
<td>.91</td>
<td>.96</td>
<td>-.82</td>
<td>-.81</td>
</tr>
<tr>
<td>.50</td>
<td>.4</td>
<td>.91</td>
<td>.96</td>
<td>-.41</td>
<td>-.41</td>
</tr>
<tr>
<td></td>
<td>.6</td>
<td>.89</td>
<td>.96</td>
<td>-.62</td>
<td>-.60</td>
</tr>
<tr>
<td></td>
<td>.8</td>
<td>.92</td>
<td>.96</td>
<td>-.82</td>
<td>-.81</td>
</tr>
<tr>
<td>1.00</td>
<td>.4</td>
<td>.93</td>
<td>.95</td>
<td>-.43</td>
<td>-.40</td>
</tr>
<tr>
<td></td>
<td>.6</td>
<td>.90</td>
<td>.93</td>
<td>-.65</td>
<td>-.60</td>
</tr>
<tr>
<td></td>
<td>.8</td>
<td>.90</td>
<td>.94</td>
<td>-.87</td>
<td>-.80</td>
</tr>
</tbody>
</table>

NOTE: The measurement error variance is \( \sigma^2_0 \), the slope is \( b_1 \), and the number of replicated cases and the number of replicated controls both equal 25. In 1,000 simulations, the coverage rates are given for nominal 90% and 95% intervals. The mean and median of the slope estimates are also listed.

in increased participation. Further details on the motivation of the study design were discussed by Wacholder et al. (1994).

The potential questionnaire design is under consideration for a study to be done by the National Cancer Institute. The study concerns the health effects of pesticide exposure (Z); diet and cooking practices (X1) and level of physical activity (X2) are also of interest. Measuring diet and cooking practices with any degree of accuracy is difficult, expensive, and time-consuming for both investigators and study participants, accurately measuring physical activity levels can be burdensome as well. Hence the investigators wish to minimize the number of subjects for whom both diet and activity are measured, because measuring both will affect compliance and accuracy.

The pattern of missingness here is nonmonotone in the sense of Little and Rubin (1987). We will show that the prospective formulas are not necessarily asymptotically correct and that in principle a correction needs to be made.

9.2 Estimating Equations and Theory

In this case, \( J = 4 \), \( \pi_j = \pi_j(Z, D) \), and \( \delta_j = 1 \) if \( Z \) only is observed \( (j = 1) \), \( (X_1, Z) \) is observed \((j = 2) \), \( (X_1, X_2, Z) \) is observed \((j = 3) \), or the entire set \((X_1, X_2, Z) \) is observed \((j = 4) \). Wacholder et al. (1994) assumed that \((X_1, X_2, Z) \) are discrete random variables.

The prospective model is \( p(D = 1 \mid X_1, X_2, Z) = H(\theta_0^* + \theta_1 X_1 + \theta_2 X_2 + \theta_3 Z) \). The marginal distribution of \((X_1, X_2, Z) \) induced by the case-control sampling scheme before “covering up” \((X_1, X_2) \) is written as \( q(x_1, x_2, z, \theta_1) \), where we have included \( \theta_1 \) as a parameter to allow various categorical data submodels; for example, the fully saturated model or the model in which \((X_1, X_2) \) is independent of \( Z \) (Bishop, Fienberg, and Holland 1975). Thus \( \theta_1 = (\theta_{11}, \theta_{12}, \theta_{13}) \) and \( \theta = (\theta_1, \theta_2, \theta_0)^T \).

There is no requirement in this formulation that \((X_1, X_2, Z) \) be discrete. If they are not, then one must specify a model for the distribution of these random variables induced by the case-control sampling scheme.

In this case, we show in the Appendix, Section A.S, that the elements of \( T_q(\theta) \) corresponding to \( \theta_1 \) need not all be zero, so that the prospective covariance formula for \( \theta_1 \) can be asymptotically conservative. When \((X_1, X_2, Z) \) are all binary and their distribution is left unspecified, it can be shown that prospective covariance formula are correct, in accordance with the Appendix, Section A.1, but the prospective covariance formula is conservative when a logistic model applies.

10. TWO-STAGE STUDIES

Our estimating equation approach can be applied even when the missing-data indicators \( \bar{d}_{ij} \) are dependent. As an illustration, consider the single-stratum two-stage study of Breslow and Cain (1988), which is based on the prospective model \( p(D = 1 \mid X) = H(\theta_0^* + \theta_1 X) \). The variable \( Z \) is a categorical surrogate with \( M \) levels. The assumption is that \( Z \) is conditionally independent of \( D \) given \( X \), which might occur, for example, when \( Z \) is a categorical level of a continuous covariate \( X \). In effect, the model is a linear logistic model, where the coefficient for \( Z \) is known to be zero.

At the first stage, we observe \((D, Z) \), we note the number of observations in each \((D, Z) \) category and then within each category select a further subsample of fixed size in which \( X \) is also observed.

In the Appendix, Section A.9, we show how to apply our estimating equation approach to rederive the Breslow and Cain result. (For a general discussion of two-stage designs, see Zhao and Lipsitz 1992.)

11. DISCUSSION

We have proposed a method for the analysis of prospective estimating equations in case-control studies. The major conclusions are that prospectively unbiased estimating equations are typically retrospectively unbiased and that the use of prospectively derived standard error estimates is asymptotically at worst conservative.

The examples we have considered allowed for multiplicative and linear logistic models, missing data, mismeasured data, and robust estimation. The techniques are applicable in general, and should prove useful in the consideration of other complex problems.

APPENDIX: TECHNICAL PROOFS

A.1 Semiparametric Perspective

Some insight into when the prospective standard errors are asymptotically correct can be gained by the following informal semiparametric argument. Suppose that the distributions of \((Z, X) \) and \((W \mid Z, X, D) \) are not parameterized. Then they are completely unrestricted, or (semiparametrically) governed by infinite-dimensional parameters \((\rho_1, \rho_2) \). The prospective likelihood can then be written as

\[ f_{X}(x, x \mid \rho_2) f_{D}(D 
mid z, x, \theta, \rho_1) f_{W}(w \mid z, x, d, \rho_2). \]

From Prentice and Pyke (1979), the prospective likelihood also can be written as

\[ f_{X}(x, x \mid \rho_2) f_{D}(D 
mid z, x, \theta, \rho_1) f_{D}(D 
mid z, x, \theta, \rho_1) f_{W}(w \mid z, x, d, \rho_2), \]

where \((\rho_1, \rho_2) \) are unrestricted (infinite-dimensional) and \( \theta_1 \) is characterized by the log-odds ratio. Because \((\rho_1, \rho_2, \rho_3) \) are all unrestricted, \((D_1, \ldots, D_n) \) is retrospectively ancillary for \( \theta_1 \), and hence the distributions of estimates of \( \theta_1 \) should be the same prospectively.
and retrospectively, even with missing $X$'s. The same argument applies even when the distribution of $(W|Z,X,D)$ is parameterized.

This informal argument is complementary to the results in Sections 5 and 6. It applies in Section 7 when the roles of $X$ and $W$ are interchanged. The result in Section 8 is not easily categorized. For the partial nuisance design of Section 9, when the distribution of $(Z,X)$ is not parameterized, the semiparametric argument also applies.

A.2 Retrospective Unbiasedness of Prospective Estimating Equations

We have no proof that prospective estimating equations are always retrospectively unbiased. But this is the case in every example we have examined, including the ones in this article. The following informal argument shows that retrospective unbiasedness is the rule, rather than the exception. This argument is a precise formulation of the well-known fact that in a classical study, if we haphazardly "sampled" from a case-control study, case or control status would follow a logistic model.

We first show what it means for the estimating equation to be retrospectively unbiased. Define $l_i(\cdot, \theta) = H_{S_i}(\cdot) [1 - H_{S_i}(\cdot)]^{-1} f_i(w|z,x,d,s,\theta_i) q_i(\cdot)$, where as before $q_i(\cdot)$ is the marginal density or mass function of $(Z,X)$ in the case-control sampling scheme. The notation $d_{\mu}(\cdot)$ means integration or summation with respect to the arguments of $d_{\mu}(\cdot).$ Then, by (10), the estimating equation is retrospectively unbiased if

$$0 = \sum_{i=1}^{\infty} \sum_{d=0}^{s} n_{ad} \int \pi_s(\cdot) \Psi_j(\cdot, \theta) l_i(\cdot, \theta) d_{\mu}(z,x,w). \quad (A.1)$$

It will be useful in later work to note that the retrospective expectation (12) is given by

$$\sum_{i=1}^{\infty} \sum_{d=0}^{s} n_{ad} E[L_{ad}(\theta)|D_{ad} = d, s] = \sum_{i=1}^{\infty} \sum_{d=0}^{s} n_{ad} q_{ad}, \quad (A.2)$$

Strictly speaking, retrospective unbiasedness of the estimating equation means that (20) holds for all $\theta$ and $q_i(\cdot)$ in an appropriate class.

Now turn to the prospective formulation. For a prospective model, the likelihood of $(D,X,Z,W)$ given $S = s$ is $l_{S_1}(\cdot, \theta^*) = q_1(z, x) H_{S_1}(\cdot) [1 - H_{S_1}(\cdot)]^{-1} f_1(w|z,x,\theta_1), \theta^* = (\theta_1, \theta_2, \ldots, \theta_D, \theta_{S_1}, \theta_{S_2}, \theta_{S_3}, \ldots, \theta_0)$, where $H_{S_1}(\cdot)$ is the marginal of $(Z,X)$ in the prospective sampling distribution in the $S_1$ stratum, and $H_{S_1}(\cdot)$ is the same as $H_{S_1}(\cdot)$ but with prospective stratum-specific intercepts. Thus prospective unbiasedness means that for all $\theta^*$ and $q_1(\cdot)$ in an appropriate class,

$$0 = \sum_{i=1}^{\infty} \sum_{d=0}^{s} n_{ad} \int \pi_s(\cdot) \Psi_j(\cdot, \theta^*) l_{S_1}(\cdot, \theta^*) d_{\mu}(z,x,w). \quad (A.3)$$

Note the similarity between (A.1) and (A.3). The equations are formally identical, with the only difference one of notation. Hence we can expect that prospectively unbiased estimating equations will also be retrospectively unbiased. In all the cases we have examined, the relationship between (A.1) and (A.3) trivially leads to retrospective unbiasedness of the estimating equation.

A.3 Sketch of Proof of the Main Theorem

Consider the retrospective formulation, where the parameter is $\theta$. By a Taylor series expansion, $\eta^{1/2}(\theta - \theta) \approx -\{T_\theta^*(\theta)\}^{-1}$ $\eta^{1/2}T_\theta(\theta)$. By a calculation similar to (A.2), $T_\theta^*(\theta)$ has expectation (13); suppressing the dependence on sample sizes, denote the result by $T_\theta^*(\theta)$.

We next compute $\text{cov} \{ n^{1/2} T_\theta(\theta) \}$ (conditioned on all the $D$'s of course). Let the notation $\{ \cdots \}$ indicate a repeat of the preceding term. Using (12), we have

$$\text{cov} \{ n^{1/2} T_\theta(\theta) \} = n^{-1} \sum_{i=1}^{\infty} \sum_{d=0}^{s} \sum_{j=1}^{r} E \{ [L_{ad}(\theta) - E[L_{ad}(\theta)|D_{ad}, s]] \{ \cdots \} | D_{ad}, s \}$$

$$= n^{-1} \sum_{i=1}^{\infty} \sum_{d=0}^{s} \sum_{j=1}^{r} E \{ L_{ad}(\theta) L_j(\theta) \} | D_{ad}, s \}$$

$$= n^{-1} \sum_{d=0}^{s} \sum_{j=1}^{r} E \{ [L_{ad}(\theta) | D_{ad}, s] \{ \cdots \} \}$$

$$= C(\theta) - \sum_{d=0}^{s} \sum_{j=1}^{r} (n_{ad}/n) E \{ [L_{ad}(\theta) | D_{ad} = d, s] \{ \cdots \} \}.$$}

It is easily seen that $E \{ L_{ad}(\theta) | D_{ad} = d, s \} = (n_{ad}/n) q_{ad}$, thus showing that

$$\text{cov} \{ n^{1/2} T_\theta(\theta) \} = C(\theta) - \sum_{d=0}^{s} \sum_{j=1}^{r} (n_{ad}/n) E \{ L_{ad}(\theta) | D_{ad} = d, s \} \{ \cdots \}.$$}

Now, using (10), we have

$$C(\theta) = \sum_{i=1}^{\infty} \sum_{d=0}^{s} (n_{ad}/n) E \{ L_{ad}(\theta) L_j(\theta) \} | D_{ad} = d, s \}$$

$$= \sum_{d=0}^{s} \sum_{j=1}^{r} (n_{ad}/n) \sum_{i=1}^{\infty} \int \pi_s(\cdot) \Psi_j(\cdot, \theta) \Psi_j(\cdot, \theta) l_{ad}(\cdot, \theta)$$

$$\times d_{\mu}(z,x,w). \quad (A.3)$$

This is identical to (14), as required.

A.4 Theory for the Classical Model

We have defined conditional unbiasedness to mean that (18) holds for all $\theta$. We have that $\theta = (\beta_0, \beta_1)^T, H(x, \theta) = H(\beta_0 + R(\beta_1, x)), \text{ and } H^{11}(x, \theta) - H(x, \theta)(1 - H(x, \theta)).$ Then conditional unbiasedness means that for any $(x, \theta),$

$$0 = \sum_{d=0}^{s} \psi(d, x, \theta) H^d(x, \theta) [1 - H(x, \theta)]^{d-1}. \quad (A.6)$$

In our notation, $n_{ad} = \int \psi(d, x, \theta) H^d(x, \theta) [1 - H(x, \theta)]^{d-1} q(x) \, dx$, so that $n_{ad} = \delta(0, \theta)$ by (A.5), and hence prospective estimating equations that are conditionally unbiased are also conditionally unbiased. It also follows that

$$T_\theta(\theta) = \int \sum_{d=0}^{s} (2d - 1) \psi(d, x, \theta) H^d(x, \theta) [1 - H(x, \theta)] \, q(x) \, dx. \quad (A.7)$$

Differentiating the right side of (A.6) with respect to $\theta$ and then integrating with respect to $q(x) \, dx$, we find that the first column of $-T_\theta(\theta)$ is

$$\int \sum_{d=0}^{s} (2d - 1) \psi(d, x, \theta) H^d(x, \theta) [1 - H(x, \theta)] \, q(x) \, dx. \quad (A.7)$$

It follows then that $\kappa_1$ equals the first column (A.7) of $-T_\theta(\theta)$ if

$$\psi(1, x, \theta) - [1 - H(x, \theta)] \{ \psi(1, x, \theta) - \psi(0, x, \theta) \},$$

which follows directly from (A.6). Based on the lemma in Section 4, we have thus shown that prospectively defined standard errors for slope parameters are retro-
respectively asymptotically correct for the general class of conditionally unbiased prospective estimating equations.

A.5 Theory for Differential Error

We first briefly sketch an argument showing that the estimating equations of Section 6 are retrospectively unbiased. Because there is only a single stratum, we drop the stratum indicators and, as in Section 6, write $k_\alpha = \{k_\alpha^{(1)}(\cdot)\}^{1}, k_\alpha^{(2)}(x)\}^{1}$. We will show that $k_\alpha^{(1)}(x) + k_\alpha^{(2)}(x) = 0$, the other cases being similar. By definition,

$$\sum_{\alpha} k_{\alpha}^{(1)}(x)$$

$$= \sum_{\alpha} \int_{0}^{1} \varpi(x, z, w)H^{(\cdot)}(\cdot) \{1 - H^{(\cdot)}\}^{1-d} \times f(w|x, z, x, \theta) q(z, x)$$

$$\times \left[ \varpi(x, z, x, w, \theta) - \frac{\varpi(x, z, w, \pi x, \theta)}{\varpi(x, z, x, \pi, \theta)} \right] \, du(z, x, w)$$

$$= \int \left[ \frac{\varpi(\cdot, \pi x, \theta)}{\varpi(\cdot, \pi, \theta)} \sum_{\alpha} \varpi(x, z, w, \theta) \{1 - H^{(\cdot)}\}^{1-d} \varpi(x, z, w, \theta) q(z, x, w) \right] \varpi(\cdot, \pi x, \theta)$$

$$= \sum_{\alpha} \int \varpi(\cdot, \pi x, \theta) \varpi(x, z, d - H, \theta) \, du(z, x, w) \frac{\varpi(\cdot, \pi x, \theta)}{\varpi(\cdot, \pi, \theta)} q(z, x, w) \left( \frac{\varpi(x, z, w, \pi x, \theta)}{\varpi(x, z, x, \pi, \theta)} \right)$$

Because this integral is zero, this yields the desired result.

We now show that this integral for the intercept, prospective covariance formulas are asymptotically correct. To do this, we must show that $k_\alpha$ is proportional to the first column of $T_\alpha(\theta)$. Let $H^{(\cdot)}(\cdot) = H^{(\cdot)}(\cdot) - H^{(\cdot)}(\cdot) \{1 - H^{(\cdot)}\}$ be the derivative of $H^{(\cdot)}$. Because $\varpi(\cdot, \pi x, \theta) = M(x, z) G(\cdot, \xi, \theta)$, direct calculations indicate that $k_\alpha = \{k_\alpha(\cdot) + k_\alpha(\cdot)\}^{1}, k_\alpha(\cdot)\}^{1}$, where

$$k_\alpha(\cdot) = \int \varpi(\cdot, \pi x, \theta) M(x, z) f(w|x, z, x, \theta) q(z, x)$$

$$\times \left[ \frac{H^{(\cdot)}(\cdot) - H^{(\cdot)}(\cdot) \{1 - H^{(\cdot)}\}}{\varpi(\cdot, \pi, \theta)} \right] \, du(z, x, w),$$

and

$$k_\alpha(\cdot) = \int \varpi(\cdot, \pi x, \theta) M(x, z) f(w|x, z, x, \theta) q(z, x)$$

$$\times \left[ \varpi(x, z, x, w, \theta) - \frac{\varpi(x, z, w, \pi x, \theta)}{\varpi(x, z, x, \pi, \theta)} \right] \, du(z, x, w),$$

With $\theta_0$ being the intercept, $(\partial/\partial \theta_0) \psi_0 = 0$, $(\partial/\partial \theta_0) \psi = M(z, x) H^{(\cdot)}(\cdot)$, $(\partial/\partial \theta_0) H^{(\cdot)}(\cdot) \{1 - H^{(\cdot)}\}^{1-d} = (2d - d - H^{(\cdot)})$, and for any function $\varpi(\cdot, \pi x, \theta), \varpi(x, z, \theta) = \varpi(\cdot, \xi, \theta) + \varpi(\cdot, \pi x, \theta)$. Writing $\varpi(\cdot, \pi x, \theta) = \varpi(\cdot, \xi, \theta)$, by direct but tedious algebra, we find that if $\Psi = \{\psi_\alpha^{(1)}, \psi_\alpha^{(2)}\}$, then

$$\varpi(\cdot, \pi x, \theta) \varpi(x, z, d - H, \theta) \times du(z, x, w) = -M(z, x) \frac{\varpi(x, z, w, \pi x, \theta)}{\varpi(x, z, x, \pi, \theta)} q(z, x, w) \left( \frac{\varpi(x, z, w, \pi x, \theta)}{\varpi(x, z, x, \pi, \theta)} \right)$$

$$= -M(z, x) \left[ \frac{\varpi(x, z, d - H, \theta)}{\varpi(x, z, \pi, \theta)} - \frac{\varpi(x, z, w, \pi x, \theta)}{\varpi(x, z, x, \pi, \theta)} \right] \frac{\varpi(x, z, w, \pi x, \theta)}{\varpi(x, z, x, \pi, \theta)} q(z, x, w) \left( \frac{\varpi(x, z, w, \pi x, \theta)}{\varpi(x, z, x, \pi, \theta)} \right).$$

(A.8)

Clearly, $(\partial/\partial \theta_0) \psi_0$ does not depend on $d$. Remembering that $(\partial/\partial \theta_0) \psi_0 - 0$, the part of the first column of $T_\alpha(\theta)$ corresponding to $\psi_0$ is

$$\sum_{\alpha} \int \varpi(\cdot, \pi x, \theta) \varpi(x, z, d - H, \theta) \times du(z, x, w) \frac{\varpi(x, z, w, \pi x, \theta)}{\varpi(x, z, x, \pi, \theta)} q(z, x, w) \left( \frac{\varpi(x, z, w, \pi x, \theta)}{\varpi(x, z, x, \pi, \theta)} \right)$$

$$= \int \left[ \frac{\varpi(x, z, w, \pi x, \theta)}{\varpi(x, z, x, \pi, \theta)} \right] q(z, x, w) \left( \frac{\varpi(x, z, w, \pi x, \theta)}{\varpi(x, z, x, \pi, \theta)} \right).$$

(A.9)

We now substitute the right side of (A.8) into (A.9), noting that it factors naturally into components depending on $\pi$ and $\phi$, the latter through $M$. We thus rewrite (A.9) as

$$-\int \left[ \frac{\varpi(x, z, w, \pi x, \theta)}{\varpi(x, z, x, \pi, \theta)} \right] q(z, x, w) \left( \frac{\varpi(x, z, w, \pi x, \theta)}{\varpi(x, z, x, \pi, \theta)} \right).$$

(A.10)

Direct but tedious algebra shows that the integrands corresponding to $\pi$ and $\phi$ in (A.10) exactly equal the integrands in $k_\alpha(\cdot)$ and $k_\alpha(\cdot)$.

Similarly, the part of the first column of $T_\alpha(\theta)$ corresponding to $\psi_\beta$s is

$$\int \left[ \frac{\varpi(x, z, \theta)}{\varpi(x, \pi, \theta)} \right] q(z, x, w) \left( \frac{\varpi(x, z, w, \pi x, \theta)}{\varpi(x, z, x, \pi, \theta)} \right).$$

(A.11)

It can be shown that the integrand of this expression equals the integrand of $k_\alpha(\cdot)$.

We have thus shown that $k_\alpha$ is proportional to the first column of $T_\alpha(\theta)$, and hence that prospective covariance formulas may be used.

A.6 Theory for Non-differential Measurement Error

The analysis requires a small notational change, namely to inter-change the roles of $x$ and $w$. Dropping the stratum indicators, (10) then becomes

$$(n/n_{\alpha}) q(z, x) \varpi(z, x) H^{(\cdot)}(\cdot) \{1 - H^{(\cdot)}\} \times \varpi(z, x, w, \theta) = \varpi(z, x, x, \xi, \theta).$$

(A.12)

In our theory, $H^{(\cdot)}(\cdot)$ replaces $H^{(\cdot)}(\cdot)$ and $\varpi(z, w, \theta, \cdot)$ replaces $\varpi(z, x, \cdot, x, \xi, \theta)$.

Dropping stratum indicators, the estimating equations for maximum likelihood then fit into our notation with $J = 2$:

$$\Psi(\cdot) = \{M(\cdot)(D - H(\cdot)), S_1, S_1\}$$

and

$$\Psi(\cdot) = \{M(\cdot)(D - H(\cdot)), S_1, S_1\},$$

where $S_1 = (\partial/\partial \theta_0) \log \{\varpi(z, w, \pi x, \theta, \cdot) \}$ and $S_1$ is defined similarly. In addition, $M(x, z) = \{1, z, \theta, u_{\alpha}, U_{\alpha}\}$, where $U_{\alpha} = (\partial/\partial \theta_0) \log \{\varpi(z, W, \theta, \cdot) \}$ and similarly for $u_{\alpha}$.

It is easy to show that these prospective estimating equations are retrospectively unbiased. With the redefinition of $k(\cdot)$, the first column of $T_\alpha(\theta)$ is

$$T_{\alpha}(\theta) = -\left[ \sum_{\alpha} \int M(z, x) H^{(\cdot)}(\cdot) \times du(z, x, w, \theta), 0^t, 0^t \right]$$

$$= -\left[ \sum_{\alpha} \int M(z, x) H^{(\cdot)}(\cdot) \times du(z, x, w, \theta), 0^t, 0^t \right]$$

(A.11)
the last step following because the only term depending on \( \alpha \) is \( f_{\theta \iota} g_{\iota}(x, \gamma, w, \delta, \theta, \iota) \), which is a density and hence integrates to 1. That (A.11) equals \( \kappa_{\alpha} \) is immediate.

We have thus shown that the Satten and Kupper (1993) "unconditional" method for prospective studies can be applied without change to retrospective studies.

### A.7 Theory for Corrections for Attenuation

Let the mean of \( W_1 \) among the controls be \( \mu_1 \) and the mean of \( W_1 - W_1 \) among the controls be \( \mu_2 = 0 \). For technical reasons having to do with the fact that \( \sigma^2_1 \) is estimated by the sample variance \( \hat{\sigma}^2_1 \), we must include an estimation equation for \( \mu_2 \) even though it is known; this estimation equation has no effect on the standard error estimates. The estimating equations for this algorithm are, with \( \theta = (\mu_1, \mu_2, \sigma^2_1, \sigma^2_2) \),

\[
\Psi_{ii} = \begin{cases}
1 & [d - H(\theta_0, \theta_1, \theta_2, \theta_3)]
\end{cases} \begin{cases}
1 & [d - H(\theta_0, \theta_1, \theta_2, \theta_3)]
\end{cases}
\]

and

\[
\Psi_{1i1} = \begin{cases}
0 & [d - H(\theta_0, \theta_1, \theta_2, \theta_3)]
\end{cases} \begin{cases}
0 & [d - H(\theta_0, \theta_1, \theta_2, \theta_3)]
\end{cases}
\]

By treating the approximations we have made to be exact, it is easily shown that the estimating equations are unbiased. If \( f(w|x) \) is the density function of \( W = (W_1, W_2) \) given \( X \), because \( \kappa_{\alpha} \) is based on the case \( d = 1 \) it follows trivially that

\[
k_{ii} = \int \{ a, \theta, g_{1}(x), 0, 0, 0, 0 \} \cdot [H_{12}(\theta)] \cdot q(x) \cdot f(w|x) \cdot d\mu(x, w),
\]

where \((a, b) = (1, 0) \) and \((0, 1) \) for \( s = 1, 2 \). It is also easily verified that the first column of \( T_{\text{q}}(\theta) \), which is based on the expectations of the derivative of \( \Psi_{ii} \) with respect to \( \theta_1, \theta_2, \theta_3 \), whereas the second column of \( T_{\text{q}}(\theta) \) is \(-k_{ii} \). Hence, subject to the levels of approximation described, the conclusion is that prospective covariance formulas may be used for the estimate of \( \theta_1 \).

We estimated standard errors using a prospective method. The variances of the terms \( \Psi_{ii} \) and \( \Psi_{1i1} \) were originally computed using the "model-free" method, with the usual exception that in the upper 3 x 3 matrix corresponding to the logistic parameters, we replaced the model-free terms by the usual information contributions.

For the terms corresponding to the derivatives of \( \Psi_{ii} \) and \( \Psi_{1i1} \), we again started with the model-free method and again modified the upper 3 x 3 matrix by substituting information contributions. For the other terms in the first three rows of \( \Psi_{ii} \) and \( \Psi_{1i1} \), we explicitly used the prospective result that eliminates the contributions of the derivatives of the terms \( g_1 \) and \( g_2 \) when they are outside of \( H(\cdot) \), because prospectively terms such as

\[
\left\langle \frac{\partial}{\partial \sigma^2_1} g_1, (W_1, \sigma^2_1, \sigma^2_2, \mu_2) \right\rangle \cdot [D - H(\theta_0, \theta_1, \theta_2, \theta_3)]
\]

have mean zero. Generalization to vector predictors is immediate.

### A.8 Theory for Partial Questionnaires

We will use the estimating equations from prospective maximum likelihood, which can be described as follows. Define

\[
G_{1} = H_{1}, \quad G_{2} = \int H q \cdot d\mu(x),
\]

\[
A_{4} = q, \quad A_{5} = \int q \cdot d\mu(x),
\]

\[
B_{j}(\cdot) = \{ \delta/\partial(\theta, \theta_0, \theta_1, \theta_2, \theta_3) \} [G_{j}(\cdot)],
\]

\[
L_{j}(\cdot) = \{ \delta/\partial \theta_0 \} G_{j}(\cdot),
\]

\[
M_{j}(\cdot) = \{ \delta/\partial \theta_1 \} A_{j}(\cdot), \quad M(\cdot) = \{ \theta, x, y, \theta \}.
\]

The prospective estimating equations are in our general form with

\[
\Psi_{ii} = \begin{cases}
0 & [d - H(\theta_0, \theta_1, \theta_2, \theta_3)]
\end{cases} \begin{cases}
0 & [d - H(\theta_0, \theta_1, \theta_2, \theta_3)]
\end{cases}
\]

By treating the approximations we have made to be exact, it is easily shown that the estimating equations are unbiased. If \( f(w|x) \) is the density function of \( W = (W_1, W_2) \) given \( X \), because \( \kappa_{\alpha} \) is based on the case \( d = 1 \) it follows trivially that

\[
k_{ii} = \int \{ a, \theta, g_{1}(x), 0, 0, 0, 0 \} \cdot [H_{12}(\theta)] \cdot q(x) \cdot f(w|x) \cdot d\mu(x, w),
\]

where \((a, b) = (1, 0) \) and \((0, 1) \) for \( s = 1, 2 \). It is also easily verified that the first column of \( T_{\text{q}}(\theta) \), which is based on the expectations of the derivative of \( \Psi_{ii} \) with respect to \( \theta_1, \theta_2, \theta_3 \), whereas the second column of \( T_{\text{q}}(\theta) \) is \(-k_{ii} \). Hence, subject to the levels of approximation described, the conclusion is that prospective covariance formulas may be used for the estimate of \( \theta_1 \).

We estimated standard errors using a prospective method. The variances of the terms \( \Psi_{ii} \) and \( \Psi_{1i1} \) were originally computed using the "model-free" method, with the usual exception that in the upper 3 x 3 matrix corresponding to the logistic parameters, we replaced the model-free terms by the usual information contributions.

For the terms corresponding to the derivatives of \( \Psi_{ii} \) and \( \Psi_{1i1} \), we again started with the model-free method and again modified the upper 3 x 3 matrix by substituting information contributions. For the other terms in the first three rows of \( \Psi_{ii} \) and \( \Psi_{1i1} \), we explicitly used the prospective result that eliminates the contributions of the derivatives of the terms \( g_1 \) and \( g_2 \) when they are outside of \( H(\cdot) \), because prospectively terms such as

\[
\left\langle \frac{\partial}{\partial \sigma^2_1} g_1, (W_1, \sigma^2_1, \sigma^2_2, \mu_2) \right\rangle \cdot [D - H(\theta_0, \theta_1, \theta_2, \theta_3)]
\]

have mean zero. Generalization to vector predictors is immediate.
\[ C(\Theta) = \sum_{z=1}^{4} \int \pi(z, \lambda) \begin{pmatrix} \Theta(z) \\ \Phi(z) \end{pmatrix} \begin{pmatrix} \Phi(z) \end{pmatrix} \begin{pmatrix} G(z) \end{pmatrix}^{-1} d\mu(z) \]
\[ + \sum_{z=1}^{4} \int \pi(z, \lambda) \begin{pmatrix} -\Theta(z) \\ \Phi(z) \end{pmatrix} \begin{pmatrix} \Phi(z) \end{pmatrix} \begin{pmatrix} G(z) \end{pmatrix}^{-1} d\mu(z). \]

When \((Z, X_1, X_2)\) are all binary and their distribution is left unspecified, detailed considerations show that prospective covariance formulas are asymptotically correct.

**A 9 Theory for Two-Stage Studies**

Let \(n_m\) be the number of observations with \(D = d\), let \(n_{md}\) be the random number of observations with \((D = d, Z = m)\), and let \(n_{m*}\) be the fixed number of observations in the second stage within each \((D, Z)\) category. Define \(\theta_{d,z} = \log(\pi_{d,z}/\pi_{d,z})\). Note that \((n_{1d}, \ldots, n_{1d})\) is a multinomial random variable with probabilities \((\theta_{2d}, \ldots, \theta_{2d})\).

Because there is only a single stratum, we will drop the stratum assignment indicators. In (10), \(j = 1\) refers to observations selected in the first stage sample, and \(j = 2\) denotes those which are not so selected. Define \(\psi_{j1}(\cdot, \Theta) = 0\) and \(\psi_{j2}(\cdot, \Theta) = (1, X') \begin{pmatrix} d - H(z) + \theta(z) \end{pmatrix} X + \log(n/n) + \log(n_{2d}/n_{2d}) \). \(\text{(A.12)}\)

Let \(\psi_{j1} = \psi_{j2} = (\cdot, \Theta)\) be the vector of size \(2M\) whose \((d M + m)\)th element equals \(I(D = d) I(Z = m) - \theta_{d,z}\).

Let \(\psi_{1} = (\psi_{11}, \psi_{12})\). Denote the logistic argument in (12) by \(H_{*}(Z, \Theta)\) and write \(H(x) = H_{*}(1, \Theta)\).

At the end of this section, the estimating equation is shown to be unbiased, the particular method being to condition on all \(n_{md}\) or, equivalently, on all the \(z\)'s. In addition, with \((D, Z)\) denoting the collection of all \((D, Z)\)'s, we later show that

\[ 0 = E \left[ n^{-1} \sum_{i=1}^{n} \delta_{i} \psi_{1}(\cdot, \Theta) \right] | \hat{D}, \hat{Z}. \] \(\text{(A.13)}\)

Next we show that the estimating equations for \((\theta_0, \theta_1)\) are uncorrelated with those for \((\theta_2, \ldots, \theta_{2d})\), so that the covariance matrix \(A(\Theta) = C(\Theta) - D(\Theta)\) is block diagonal. To see this, first note that the off-diagonal term in the covariance matrix is

\[ \sum_{i=1}^{n} \delta_{i} \psi_{2}(\cdot, \Theta) \]

\[ \times \left[ \sum_{i=1}^{n} \delta_{i} \psi_{2}(\cdot, \Theta) + \sum_{i=1}^{n} \delta_{i} \psi_{2}(\cdot, \Theta) \right] | \hat{D}. \]

The terms associated with \(\psi_{2} = \psi_{22}\) depend only on \((D, Z)\). So, if we condition on this term and apply (A.13), then we have the desired result.

Breslow and Cain did not use our estimating equation approach, but their results are equivalent to ours, except that they worked with the parameterization \(\pi_{md} = \log(\pi_{md})\). In our notation we have shown that

\[ T_{d}(\Theta) = \begin{bmatrix} T_{d1} & T_{d13} \\ T_{d13} & T_{d23} \end{bmatrix}; \quad A(\Theta) = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}. \]

Except for the change in parameterization and the fact that our asymptotics are based on the total sample size \(n\) rather than on the total \(n = \sum_{d,z} n_{md}\) (a notational difference of no effect on the final results), their term \(H \) corresponds to our \(T_{d1} \), their term \(A \) to our \(T_{d2} \), their term \(B \) to our \(A_{d2} \), and, as we show at the end of this section, their term \(G \) to our \(A_{d1} \).

We now turn to filling in the main technical steps. We first show (A.13). The notation is that \(H_{*}(\Theta, \Theta)\) refers to the logistic argument in (A.12). Because the conditional density or probability mass function \(f(x | z, d) = f(x | z, d) / \theta_{2d}\), then

\[ E \left[ \sum_{i=1}^{n} \delta_{i} \psi_{1}(\cdot, \Theta) | \hat{D}, \hat{Z}, \hat{d} \right] \]

\[ = \sum_{m=1}^{M} \sum_{d=0}^{D} n_{md} E \left[ \psi_{1}(\cdot, \Theta) | \hat{D}, \hat{Z}, \hat{d} \right] \]

\[ = \sum_{m=1}^{M} \sum_{d=0}^{D} n_{md} E \left[ \psi_{1}(\cdot, \Theta) | \hat{D}, \hat{Z} \right] \]

\[ = \sum_{m=1}^{M} \sum_{d=0}^{D} n_{md} \psi_{1}(\cdot, \Theta) \left[ \sum_{d=0}^{D} n_{md} \theta_{2d} \right] \]

\[ \times H(d | \cdot) | 1 - H(d | \cdot) | 1 - H(d | \cdot) \]

\[ \quad \cdot \mu(d | x). \]

\(\text{(A.14)}\)

However, the term in square brackets in (A.14) equals zero, proving (A.13).

Next we prove that the estimating equation is unbiased. The part corresponding to \(\psi_{1}(\cdot, \Theta)\) is unbiased by (A.13). For the other part, for specificity consider the estimating equation corresponding to \(\theta_{2d}\). This has expectation

\[ (n_1/n) \sum_{i=1}^{n} \left[ I(Z = m - \theta_{2d}) \right] H(d | \cdot) \psi_{1}(\cdot, \Theta) \psi_{1}(\cdot, \Theta) \]

\[ = \sum_{m=1}^{M} \sum_{d=0}^{D} n_{md} E \left[ \psi_{1}(\cdot, \Theta) | \hat{D}, \hat{Z}, \hat{d} \right] \]

\[ = \sum_{m=1}^{M} \sum_{d=0}^{D} n_{md} E \left[ \psi_{1}(\cdot, \Theta) | \hat{D}, \hat{Z} \right] \]

\[ = \sum_{m=1}^{M} \sum_{d=0}^{D} n_{md} \psi_{1}(\cdot, \Theta) \left[ \sum_{d=0}^{D} n_{md} \theta_{2d} \right] \]

\[ \times H(d | \cdot) | 1 - H(d | \cdot) | 1 - H(d | \cdot) \]

\[ \cdot \mu(d | x). \]

\(\text{(A.14)}\)

Because \((n_1/n) H(d | \cdot) \psi_{1}(\cdot, \Theta) \psi_{1}(\cdot, \Theta) \)

\[ = \sum_{m=1}^{M} \sum_{d=0}^{D} n_{md} E \left[ \psi_{1}(\cdot, \Theta) | \hat{D}, \hat{Z}, \hat{d} \right] \]

\[ = \sum_{m=1}^{M} \sum_{d=0}^{D} n_{md} E \left[ \psi_{1}(\cdot, \Theta) | \hat{D}, \hat{Z} \right] \]

\[ = \sum_{m=1}^{M} \sum_{d=0}^{D} n_{md} \psi_{1}(\cdot, \Theta) \left[ \sum_{d=0}^{D} n_{md} \theta_{2d} \right] \]

\[ \times H(d | \cdot) | 1 - H(d | \cdot) | 1 - H(d | \cdot) \]

\[ \cdot \mu(d | x). \]

This last term is Breslow and Cain's matrix \(G\).
REFERENCES


