PROSPECTIVE ANALYSIS OF LOGISTIC
CASE–CONTROL STUDIES

R. J. Carroll & Suojin Wang
Department of Statistics
Texas A&M University
College Station, TX 77843

C. Y. Wang
Division of Public Health Sciences
Fred Hutchinson Cancer Research Center
Seattle, WA 98104

September 21, 1994

Abstract

In a classical case-control study, Prentice & Pyke (1979) propose to ignore the study design and instead base estimation and inference upon 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 estimators, which ignore the case-control study aspect and instead are based upon a random-sampling formulation, are typically consistent for non-intercept 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 formulae 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.

Some Key Words: Asymptotics; Case–control studies; Corrections for attenuation; Differential measurement error; Estimating equations; Missing data; Measurement error; Robust estimates.
AUTHORS' AFFILIATIONS

R. J. Carroll and Suojin Wang are professor and associate professor, respectively, in the Department of Statistics, Texas A&M University, College Station TX 77843-3143. C. Y. Wang is assistant member, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98104.

Carroll and C. Y. Wang’s research was supported by a grant from the National Cancer Institute (CA-57030). S. Wang’s research was supported by grants from the Texas Advanced Research Program (Grant 160802) and the National Science Foundation (DMS-9200610). We are grateful to Mitch Gail and Sholom Wacholder, whose questions led to the general formulation of this paper, and to Jamie Robins for help with the argument in section 12.1. We also wish to thank two referees for their comments, and an associate editor who read the paper in painstaking detail and suggested many useful changes.
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 ascertained, along with the values of covariates $(Z, X)$, these being related via the logistic regression model

$$\text{pr}(D = 1 | Z, X) = H(\theta_0^* + \theta_1^* Z + \theta_2^* X),$$

(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 one samples from within the cases and controls to observe the covariates. The analysis of case–control studies of this type was described by Prentice & 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 $(\theta_1, \theta_2)$ 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:

(i) One might replace the classical logistic regression parameter estimates by robust methods of estimation (Copas, 1988; Künsch, Stefanski & Carroll, 1989; Carroll & Pederson, 1993).

(ii) When $X$ is measured with error, there is a large literature dealing with techniques for measurement error corrections in logistic regression (e.g., Stefanski & Carroll, 1987; Rosner, Willett & Spiegelman, 1989; Satten & Kupper, 1993; Carroll & Stefanski, 1994).

(iii) In problems with partially missing data, one can use likelihood techniques (Little & Rubin, 1987) or unbiased estimating equations due to Robins, Hsieh & Newey (1994).

While 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 so, 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, i.e., 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 (section 12.1), we sketch an informal argument derived from a semiparamet-
ric perspective that suggests that prospectively computed standard errors are retrospectively correct whenever the distribution of \((Z, X)\) is left unrestricted. Much of this paper is a formalization of this argument, along with consideration of cases which 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 & Wacholder (1993).

Here is an outline of the paper. In section 2, we review the known results on estimating equations, estimating functions and sandwich covariance estimation in prospective studies. Using this background, we provide a simple argument showing 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. The two main results are stated in section 4.

The rest of the paper consists of consideration of 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 & Wacholder (1993) for multiplicative models and Wang & Carroll (1993, 1994) 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. In section 6, we apply 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 section 6). In section 7, we consider measurement error models with nondifferential measurement error in which a validation study can be done, using the simple prospective likelihood methods due to Satten & Kupper (1993). In section 8, we discuss measurement error models when validation is not possible, and use prospective correction-for-attenuation methods. In all three cases, prospective standard errors are asymptotically correct retrospectively.

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

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 \( \Pr(D = 1|X) = H(\theta_0^* + \theta_1 X) \). The argument uses nothing more than standard estimating equation theory; we will outline this theory and the nomenclature as we go along. Extensions to complex problems require little more than a change in notation.

2.1 Prospective Sampling

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

\[
0 = \sum_{i=1}^{n} \left( \frac{1}{X_i} \right) \{ D_i - H(\theta_0^* + \theta_1 X_i) \} = \sum_{i=1}^{n} \psi(D_i, X_i, \Theta^*). \tag{2}
\]

The entire term on the right-hand-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, i.e., that 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:

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

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

\[
n^{1/2} \left( \hat{\Theta}^* - \Theta^* \right) \approx \text{Normal} \left\{ 0, B^{-1}(\Theta^*)A(\Theta^*)B^{-1}(\Theta^*) \right\}, \text{ where} \tag{5}
\]

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

\[
A(\Theta^*) = n^{-1} \text{cov} \left\{ \sum_{i=1}^{n} \psi(D_i, X_i, \Theta^*) \right\}. \tag{7}
\]
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, i.e., use (4), to conclude that if we define

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

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, i.e., 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 & McCullagh, 1993). For example, the resulting model-free estimate of \(B(\Theta^*)\) is just

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

### 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 \(\theta_0 = \theta_0^* + \log(n_1/n_0) - \log\{\text{pr}(D=1)/\text{pr}(D=0)\}\), where \(\text{pr}(D=1)\) is the unknown prospective rate. Prentice & Pyke (1979) show that if \(\Theta = (\theta_0, \theta_1)^t\) and we replace \(\Theta^*\) by \(\Theta\), then the estimating function (2) is still unbiased, i.e., (3) holds. However, 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 rewrite (7) as follows,

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

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

\[ = C(\Theta) - D(\Theta). \]

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 basing estimation as if the correct covariance matrix were \( n^{-1} B^{-1}(\Theta) C(\Theta) B^{-t}(\Theta) \).
- However, the proper covariance is \( n^{-1} B^{-1}(\Theta) \{ C(\Theta) - D(\Theta) \} B^{-t}(\Theta) \).
- Since \( D(\Theta) \) is positive semidefinite, prospective covariance formulae are at worst conservative.

### 2.3 Further Steps

The reasoning given above 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.

The analysis given shows that prospective covariance formulae are at worst conservative, but no insight is given as to when these formulae 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 paper, 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 which may be partially missing by \( X \). Anticipating the possibility that the study may be stratified, we use the stratum assignment 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, e.g., blood pressure measured at a single time point as a proxy for long-term blood pressure. The vector of parameters of major interest is denoted by \( \theta_l \), e.g., in (1), \( \theta_l = (\theta_{1l}, \theta_{2l})^t \).

If there were no missing data, we assume a sampling mechanism of a classical case–control study within each stratum \( S = s \), with \( n_{1s} \) cases, \( n_{0s} \) controls and \( n_s = n_{0s} + n_{1s} \) observations. The total
sample size is \( n = \sum n_s \). We assume that the terms \( n_{js}/n \) 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

\[
pr(D = 1|Z, X, S = s) = H \{ \theta^*_0 + R_s(\theta_1, \theta_2, Z, X) \},
\]

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

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

\[
\frac{q_s(z, x)H^d_s(\cdot)}{H_s(\cdot)} \{1 - H_s(\cdot)\}^{1-d} f(w|z, x, d, s, \theta_2), \quad \text{where} \quad H_s(\cdot) = H \{ \theta^*_0 + R_s(\theta_1, \theta_2, z, x) \}.
\]

In (10), \( q_s(\cdot) \) is the marginal density of \((Z, X)\) in stratum \( S = s \) induced by the case-control sampling scheme, while \( \theta^*_0 = \theta^*_0 + \log(n_1/s)/\log(\text{pr}(D = 1|S = s)/\text{pr}(D = 0|S = s)) \), where \text{pr}(D = 1|S = s) is the prospective rate in stratum \( s \). We write \( \Theta = (\theta_0, \ldots, \theta_0, \theta_1, \theta_2)^t \), the retrospective parameter, and \( \Theta^* = (\theta^*_0, \ldots, \theta^*_0, \theta_1, \theta_2)^t \), 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 (see below for an example), \( \Delta = (\delta_1, \ldots, \delta_J) \) 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 \( X_{(2)} \) 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, etc.

Table 1 summarizes the notation.
<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>$\delta_j$</td>
<td>Indicator that $X$ is missing with pattern number $j$</td>
</tr>
<tr>
<td>$\pi_j(z, w, s, d)$</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>$\theta_1$</td>
<td>Non-intercept parameter in the prospective logistic model</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>Error model parameter for the distribution of $W$</td>
</tr>
</tbody>
</table>

**Table 1: Notation used in the paper.**

**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. A natural approach to estimation then is to use prospective estimating equations. Let $\delta_{ij}$ denote the value of $\delta_j$ for the $i$th individual in the $s$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

$$0 = n^{-1} \sum_{s=1}^{S} \sum_{i=1}^{n_s} \sum_{j=1}^{J} \delta_{ij} \Psi_{js}(D_{is}, Z_{is}, X_{is}, W_{is}, s, \Theta)$$

$$= n^{-1} \sum_{s=1}^{S} \sum_{i=1}^{n_s} \mathcal{L}_{is}(\Theta) = \mathcal{T}_n(\Theta).$$

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 $(Z_{is}, X_{is}, W_{is}, \Delta_{is})$ are independent and identically distributed as $i$ varies. The individual components of these vectors are of course dependent. The assumption of independent and identi-
cally 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

$$0 = \sum_{s=1}^{S} \sum_{i=1}^{n_s} E \{ L_{is} (\Theta) | D_{is}, s \}.$$  \hfill (12)

Assumption (12) is satisfied in all the cases we have examined. As described in more detail in section 12.2, this appears to be a general phenomenon, and not simply a matter of convenient example selection on our part.

In order to state the main result, we make the following definitions. Define $\ell_s (\cdot, \Theta) = H_s^d (\cdot) \{ 1 - H_s (\cdot) \}^{1-d} f(w|z, x, d, s, \theta_s) q_s (\cdot)$. The notation $d\mu (\cdot)$ 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

$$T_{\Theta} (\Theta) = \sum_{s=1}^{S} (n_s / n) \sum_{d=0}^{1} \sum_{j=1}^{J} \int \pi_j (\cdot) \Psi_{js\Theta} (\cdot, \Theta) \ell_s (\cdot, \Theta) d\mu (z, x, w);$$  \hfill (13)

$$C (\Theta) = \sum_{s=1}^{S} (n_s / n) \sum_{d=0}^{1} \sum_{j=1}^{J} \int \pi_j (\cdot) \Psi_{js\Theta} (\cdot, \Theta) \Psi_{js\Theta} (\cdot, \Theta) \ell_s (\cdot, \Theta) d\mu (z, x, w);$$  \hfill (14)

$$\kappa_{ds} = \int \sum_{j=1}^{J} \pi_j (\cdot) \Psi_{js\Theta} (\cdot, \Theta) \ell_s (\cdot, \Theta) d\mu (z, x, w).$$
**Theorem:** Let \( \hat{\Theta} \) be the solution to (11) under retrospective sampling, and let \( \hat{\Theta}^* \) be the solution under prospective sampling. Under appropriate regularity conditions,

**Retrospectively**, \( n^{1/2}(\hat{\Theta} - \Theta) \) is asymptotically normally distributed with mean zero and covariance matrix

\[
\{T_\Theta(\Theta)\}^{-1} \left[ C(\Theta) - \sum_{s=1}^{S} \sum_{d=0}^{1} \left\{ \frac{n_s^2}{(n_d s)} \right\} \kappa_{ds} \right] \{T_\Theta(\Theta)\}^{-1}.
\]

**Prospectively**, define \( \ell_s(\cdot, \Theta^*) = q_s(z, x)H_s^d(\cdot)\{1 - H_s^d(\cdot)\}^{1-d} f(\cdot | \cdot, \theta_2) \), where \( q_s(\cdot) \) is the marginal of \((Z, X)\) in the prospective sampling distribution in the \( s \)th stratum, and \( H_s^d(\cdot) \) is the same as \( H_s(\cdot) \) but with prospective stratum-specific intercepts. Let \( C_s(\Theta^*) \) and \( T_\Theta(\Theta^*) \) be defined similarly to \( C(\Theta) \) and \( T_\Theta(\Theta) \) but with \( \ell_s \) and \( \Theta \) replaced by \( \ell_s^* \) and \( \Theta^* \). Then \( n^{1/2}(\hat{\Theta}^* - \Theta^*) \) is asymptotically normally distributed with mean zero and covariance matrix

\[
\{T_\Theta^*(\Theta^*)\}^{-1} C_s(\Theta^*) \{T_\Theta^*(\Theta^*)\}^{-1}.
\]

The proof of the theorem is sketched in section 12.3.

### 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

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

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

Here we state a simple sufficient condition which guarantees that prospectively derived standard errors are asymptotically correct. For most cases, \( \kappa_0 = -\kappa_1 \) for each stratum, and we assume this here. This leads to the following simple result.

**Corollary:** Suppose that \( \kappa_1 = -\kappa_0 \) and that \( \kappa_1 \) is proportional to the \( s^{th} \) column of \( T_\Theta(\Theta) \) for \( s = 1, \ldots, S \). Then prospectively derived covariance formulae for \((\theta_1, \theta_2)\) are asymptotically correct. More generally, the result holds if the rows of \( T_\Theta(\Theta) \kappa_1 \) corresponding to \( \theta_1 \) 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 formulae (section 2), namely

\[
\left\{ T_n \left( \hat{\Theta} \right) \right\}^{-1} n^{-1} \sum_{s=1}^{S} \sum_{i=1}^{n_s} \mathcal{L}_{is} \left( \hat{\Theta} \right) \mathcal{L}_{is}^t \left( \hat{\Theta} \right) \left\{ T_n \left( \hat{\Theta} \right) \right\}^{-t},
\]

where

\[
T_n \left( \hat{\Theta} \right) = n^{-1} \sum_{s=1}^{S} \sum_{i=1}^{n_s} \frac{\partial}{\partial \Theta} \mathcal{L}_{is} \left( \hat{\Theta} \right).
\]

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 that prospective formulae are conservative, there are two ways to construct asymptotically correct covariance estimates. The preferred method is to begin with (15), and estimate \( T_\Theta \left( \Theta \right) \) and \( C \left( \Theta \right) \) by prospective formulae; 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 \( \kappa_{ds} \) in (15), use \( \hat{\kappa}_{ds} = n_{s}^{-1} \sum_{i=1}^{n_s} I(D_{is} = d) \mathcal{L}_{is} \left( \hat{\Theta} \right), \) a model-free consistent estimate.

This hybrid approach, where \( \kappa_{ds} \) is estimated without a model and \( T_\Theta \left( \Theta \right) \) and \( C \left( \Theta \right) \) typically being based on a prospective model, will work for most cases. However, 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 \( \left\{ T_n \left( \hat{\Theta} \right) \right\}^{-1} B_n \left( \hat{\Theta} \right) \left\{ T_n \left( \hat{\Theta} \right) \right\}^{-t}, \)

where

\[
B_n \left( \Theta \right) = n^{-1} \sum_{s=1}^{S} \sum_{d=\Theta}^{1} \sum_{i=1}^{n_s} I(D_{is} = d) \left[ \mathcal{L}_{is} \left( \Theta \right) - \hat{m}(d, s, \Theta) \right]^{t},
\]

where \( \hat{m}(d, s, \Theta) = n_{ds}^{-1} \sum_{i=1}^{n_s} I(D_{is} = d) \mathcal{L}_{is} \left( \Theta \right) \) is an estimate of \( E \left\{ \mathcal{L}_{is} \left( \Theta \right) | D_{is} = d \right\} \).

5 CLASSICAL STUDIES

By a classical case–control study, we mean one with no missing data and a single stratum. Dropping the subscripts \( (j,s) \) which indicate missing data pattern and stratum number, from (9) we have \( \text{pr}(D = 1 | X) = H \{ \theta_0^* + R(\theta_1, X) \}. \) 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 & Wacholder (1993) on multiplicative models and Wang & Carroll (1993, 1994) on robust estimation.
To motivate the class of estimators, first consider simple linear logistic regression with $R(x, \theta_1) = \theta_1^T x$, and recall from (2) that the estimating function for the maximum likelihood estimator is
\[ \psi(d, x, \Theta^*) = (1, x)^T \{ d - H(\theta_0^* + \theta_1^T x) \}. \]
By assumption, prospectively
\[ E \{ \psi(D, X, \Theta^*) | X \} = 0, \tag{18} \]
since $pr(D = 1 | X) = H(\theta_0^* + \theta_1^T X)$. For the prospective maximum likelihood estimator in the general model, the estimating equation for the maximum likelihood estimator is $\psi(d, x, \Theta^*) = \{ 1, (\partial / \partial \theta_1) R(\theta_1, x) \} [d - H(\theta_0^* + R(\theta_1, x))]$, and (18) still holds. The same condition applies to all the robust estimators discussed by Carroll & Peterson (1993).

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

In the appendix section 12.4, we will show the following result.

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

The result is anticipated from section 12.1, 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 one or as a t-random variable with 3 degrees of freedom. We chose $\theta_0^* = -4.0$, $\theta_1 = -0.4, -0.6, -0.8$. 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 in section 4.1 of Carroll & Pederson (1993). The results are given in Table 2. Note that in all cases, both the ordinary and the robust methods very nearly attain their nominal levels.
Table 2: Simulation of ordinary and robust logistic regression. In 500 simulations, the coverage rates are given for nominal 90% and 95% intervals. The median of the slope estimates is also listed.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Ordinary</th>
<th>Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_1$</td>
<td>90%</td>
</tr>
<tr>
<td>Normal</td>
<td>-0.4</td>
<td>.866</td>
</tr>
<tr>
<td></td>
<td>-0.6</td>
<td>.878</td>
</tr>
<tr>
<td></td>
<td>-0.8</td>
<td>.912</td>
</tr>
<tr>
<td>$t(3)$</td>
<td>-0.4</td>
<td>.912</td>
</tr>
<tr>
<td></td>
<td>-0.6</td>
<td>.888</td>
</tr>
<tr>
<td></td>
<td>-0.8</td>
<td>.906</td>
</tr>
</tbody>
</table>

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 Pepe & Fleming (1991), Carroll & Wand (1991) and Reilly & Pepe (1993).

We will say that measurement error is non-differential 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) describe a general class of prospectively unbiased estimating equations for missing and mismeasured 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 $H(\theta_0^* + \theta_1^tZ + \theta_2^tX)$. 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 $\psi(d, z, x, \Theta)$ be the usual logistic estimating function $M(z, x)\{d - H(\cdot)\}$, where $M(z, x) = (1, z^t, x^t)^t$. Write the conditional density or mass function for $W$ as $f(w|z, x, d, \Theta) = f(w|z, x, d, \theta_3)$. 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

$$G(z, x, w, \xi, \Theta) = \sum_{d=0}^{1} \xi(z, x, w, d) f(w|z, x, d, \Theta) H^d(\cdot) \{1 - H(\cdot)\}^{1-d}.$$ 

Then for an arbitrary function $\phi(d, z, w)$ (Robins, et al. show how one can choose $\phi$ prospectively, and the same method applies retrospectively), with $j = 1$ being the case that all of $(Z, X, W, D)$ are observed, we define

$$
\Psi_1(\cdot, \Theta) = \left[ \begin{array}{c}
\psi(D, Z, X, \Theta) - \frac{G(Z, X, W, \pi, \Theta)}{G(Z, X, W, \pi, \Theta)} - \frac{\xi(Z, X, W, \pi, \Theta)}{G(Z, X, W, \pi, \Theta)} \\
\chi(Z, X, W, D, \Theta) - \frac{G(Z, X, W, \pi, \Theta)}{G(Z, X, W, \pi, \Theta)}
\end{array} \right];
\Psi_2(\cdot, \Theta) = \left[ \begin{array}{c}
\phi(D, Z, W) \\
0
\end{array} \right].
$$

Note that since 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, section 12.5. In that section, we also show that prospectively derived standard errors are asymptotically correct.

7 LIKELIHOOD & NONDIFFERENTIAL MEASUREMENT ERROR

Satten & 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 & Kupper formulate the problem as follows. For all subjects, $(D, Z, W)$ is observed. However, for the $i^{th}$ individual either $X_i$ is also observed ($\delta_i = 1$) or $X_i$ is not observed ($\delta_i = 0$). If $f_{X|Z, W, D}$ is the density or mass function of $X$ given $(Z, W, D)$, the prospective likelihood can be written as

$$
\prod_{i=1}^{n} \left[ f_{X|Z, W, D}(X_i|Z_i, W_i, D_i) \{pr(D_i = 1|Z_i, W_i)\}^{D_i} \{1 - pr(D_i = 1|Z_i, W_i)\}^{1-D_i} \right].
$$

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

$$R(Z, W, \Theta) = \log \left[ E \{\exp(\theta_2 X)|Z, W, D = 0, \Theta\} \right].$$
Prospectively, they show that \( \Pr(D = 1|Z, W) = H \{ \theta_0^* + \theta_1^* Z + R(Z, W, \Theta) \} \), and further that the ratio of the density or mass functions is

\[
\frac{f_{X|Z,W,D}(x|z,w,d = 1,\Theta)}{f_{X|Z,W,D}(x|z,w,d = 0,\Theta)} = \exp \{ \theta_1^* x - R(z,w,\Theta) \},
\]

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 section 12.6, we show that maximizing this prospective likelihood leads to estimators which are retrospectively consistent and standard errors which 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 we may observe only an unbiased surrogate (defined in section 6) \( W \) for \( X \), possibly with replication on a subset of the data. If the variance of the measurement error is known or estimated from external data sources, 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 Rosner, et al. (1989), Rosner, Spiegelman & Willett (1990), Carroll & Stefanski (1990), Gleser (1990), Liu & Liang (1992) and Schafer (1993). Carroll & Stefanski (1994) describe an instrumental variables method.

These methods differ fundamentally from the moments methods of section 6 in that they apply in the common case that \( X \) is never observable, e.g., blood pressure, diet history, etc.

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 & 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 & Lubin (1993). While all of these methods can be analyzed by our general theory, in this section we will define and investigate a version of the correction for attenuation methodology which is 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 $X$, i.e., 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 $\sigma_u^2$. To keep the analysis simple we will ignore $Z$, and study the prospective model $H(\theta_0^* + \theta_1 X)$. There are two strata $(s = 1, 2)$, one in which only $W_1$ is observed $(s = 1)$, the other for which both $(W_1, W_2)$ are observed $(s = 2)$. Set $J = 1$ and $\pi_j = 1$.

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

$$\text{pr}(D = 1|W_i) \approx H \{\theta_0^* + \theta_1 m_1(W_i)\};$$

$$\text{pr}(D = 1|\overline{W}) \approx H \{\theta_0^* + \theta_1 m_2(\overline{W})\},$$

where $m_1(W_i) = E(X|W_i)$ and $m_2(\overline{W}) = E(X|\overline{W})$. 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. However, 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 $m_1(W_i) \approx E(X|W_i, D = 0)$ and $m_2(\overline{W}) \approx E(X|\overline{W}, D = 0)$, approximations which we will henceforth treat as exact. Let $\mu_w$ be the mean of $W_i$ among the controls. Following Carroll & Stefanski (1990) and Gleser (1990), for $s = 1, 2$ estimates of the best linear approximations to these regressions are

$$g_1(W_i, \sigma_u^2, \sigma_w^2, \bar{\mu}_w) = \bar{\mu}_w + \left\{(\sigma_u^2 - \sigma_w^2)/\sigma_w^2\right\} (W_i - \bar{\mu}_w);$$

$$g_2(\overline{W}, \sigma_u^2, \sigma_w^2, \bar{\mu}_w) = \bar{\mu}_w + \left\{(\sigma_u^2 - \sigma_w^2)/(\sigma_u^2 - \sigma_w^2/2)\right\} (\overline{W} - \bar{\mu}_w),$$

respectively, where $\sigma_u^2$ is the sample variance of $W_1$ among all the controls, and $\sigma_w^2$ is the sample variance of $(W_1 - W_2)/2^{1/2}$ among the replicated data.

The algorithm then is as follows. Use the replicated data to construct $\sigma_u^2$ and use the $W_1$’s from all the control data to construct $\bar{\mu}_w$ and $\sigma_w^2$. 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 section 12.7 we show that prospective covariance formulae may be used for the estimate of $\theta_1$.

The preceding analysis is readily extended to problems with vector predictors.
<table>
<thead>
<tr>
<th>$\sigma_u^2$</th>
<th>$\beta_1$</th>
<th>90%</th>
<th>95%</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>-0.4</td>
<td>0.89</td>
<td>0.96</td>
<td>-0.41</td>
<td>-0.41</td>
</tr>
<tr>
<td></td>
<td>-0.6</td>
<td>0.89</td>
<td>0.96</td>
<td>-0.61</td>
<td>-0.60</td>
</tr>
<tr>
<td></td>
<td>-0.8</td>
<td>0.91</td>
<td>0.96</td>
<td>-0.82</td>
<td>-0.81</td>
</tr>
<tr>
<td>0.50</td>
<td>-0.4</td>
<td>0.91</td>
<td>0.96</td>
<td>-0.41</td>
<td>-0.41</td>
</tr>
<tr>
<td></td>
<td>-0.6</td>
<td>0.89</td>
<td>0.95</td>
<td>-0.62</td>
<td>-0.60</td>
</tr>
<tr>
<td></td>
<td>-0.8</td>
<td>0.92</td>
<td>0.96</td>
<td>-0.82</td>
<td>-0.81</td>
</tr>
<tr>
<td>1.00</td>
<td>-0.4</td>
<td>0.93</td>
<td>0.95</td>
<td>-0.43</td>
<td>-0.40</td>
</tr>
<tr>
<td></td>
<td>-0.6</td>
<td>0.90</td>
<td>0.93</td>
<td>-0.65</td>
<td>-0.60</td>
</tr>
<tr>
<td></td>
<td>-0.8</td>
<td>0.90</td>
<td>0.94</td>
<td>-0.87</td>
<td>-0.80</td>
</tr>
</tbody>
</table>

Table 3: Simulation of correction for attenuation. The measurement error variance is $\sigma_u^2$, the slope is $\beta_1$, 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.

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 of the cases and controls was 25. Prospectively, the variable $X$ was generated as a normal random variable with mean zero and variance one. We chose $\theta_0^* = -4.0$, $\theta_1 = -0.4, -0.6, -0.8$. 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 $\sigma_u^2 = 0.25, 0.5, 1.0$, these representing small, moderate and large measurement error.

We estimated $\theta_1$ using the algorithm described above. Standard errors were computed using a prospective method described in the appendix section 12.7. The results, given in Table 3, indicate that the prospectively derived confidence intervals very nearly achieve 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 while 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 less burden on study subjects, possibly resulting in increased participation. Further details on the motivation of the study design are discussed by Wacholder, et al. (1994).

The partial 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 \( (Z) \), while diet and cooking practices \( (X_1) \) and level of physical activity \( (X_2) \) are also of interest. Measuring diet and cooking practices with any degree of accuracy is difficult, expensive and time-consuming both for 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 non-monotone in the sense of Little & Rubin (1987), and we will show that the prospective formulae 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 \( (j = 2), (X_2, Z) \) is observed \( (j = 3) \) or the entire set \( (X_1, X_2, Z) \) is observed \( (j = 4) \). Wacholder, et al. (1994) assume that \( (X_1, X_2, Z) \) are discrete random variables.

The prospective model is \( \text{pr}(D = 1|X_1, X_2, Z) = H(\theta_0^\pi + \theta_1X_1 + \theta_2X_2 + \theta_3Z) \). 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 \( q(x_1, x_2, z, \theta_2) \), where we have included \( \theta_2 \) as a parameter to allow various categorical data submodels, e.g., the fully saturated model or the model in which \( (X_1, X_2) \) is independent of \( Z \) (Bishop, Fienberg & Holland, 1975). Thus, \( \theta_1 = (\theta_{11}, \theta_{12}, \theta_{13})^t \) and \( \Theta = (\theta_0, \theta_1^\prime, \theta_2^\prime)^t \).

There is no requirement in this formulation that \( (X_1, X_2, Z) \) be discrete. If they are not, 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 section 12.8 that the elements of \( \mathcal{T}_0(\Theta)\mathbf{k}_{11} \) 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 section 12.1, 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 $\delta_{ij}$ are dependent. As an illustration, consider the single-stratum two-stage study of Breslow & Cain (1988), which is based upon the prospective model $\text{pr}(D = 1|X) = H(\theta_0^* + \theta_i^* X)$. The variable $Z$ is a categorical surrogate with $M$ levels. The assumption is that $Z$ is conditionally independent of $D$ given $X$, which might for example occur 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 $(D, Z)$, 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 12.9, we show how to apply our estimating equation approach to rederive the Breslow & Cain result. For a general discussion of two-stage designs, see Zhao & 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 (i) prospectively unbiased estimating equations are typically retrospectively unbiased; and (ii) 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.

REFERENCES


Rosner, B., Spiegelman, D. & Willett, W. C. (1990), “Correction of Logistic Regression Relative Risk Estimates and Confidence Intervals for Measurement Error: the Case of Multiple Covari-


### 12 APPENDIX

#### 12.1 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 \((W|Z, X, D)\) 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

\[
f_{Z, X}(z, x | \rho_1) f_{D|Z, X}(d|z, x, \theta_0, \theta_1) f_{W|Z, X, D}(w|z, d, \rho_2).
\]

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

\[
f_{Z, X|D}(z, x | d, \theta_1, \rho_3) f_{D}(d, \rho_4) f_{W|Z, X, D}(w|z, x, d, \rho_2),
\]

where \((\rho_3, \rho_4)\) are unrestricted (infinite dimensional) and \(\rho_1\) is characterized by the log-odds ratio. Since \((\rho_2, \rho_3, \rho_4)\) are all unrestricted, \((D_1, ..., D_n)\) 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 section 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 questionnaire design of section 9, when the distribution of \((Z, X)\) is not parameterized, the semiparametric argument also applies.

12.2 Retrospective Unbiasedness of Prospective Estimating Equations

We have no proof that prospective estimating equations are always retrospectively unbiased. However, this is the case in every example we have examined, including the ones in the paper. The following informal argument shows that retrospective unbiasedness is the rule, rather than the exception. This argument is a precise manifestation of the well-known fact that in a classical study, if we fictitiously "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 
\[
\ell_s(\cdot, \Theta) = H_s^2(\cdot) \{1 - H_s(\cdot)\}^{1-d} f(w|z, x, d, s, \theta_2) q_s(\cdot),
\]
where as before \(q_s(\cdot)\) 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 \(\mu(\cdot)\). Then, by (10), the estimating equation is retrospectively unbiased if
\[
0 = \sum_{s=1}^{S} \sum_{d=0}^{1} \sum_{j=1}^{J} n_s \int \pi_j(\cdot) \Psi_{js}(\cdot, \Theta) \ell_s(\cdot, \Theta) d\mu(z, x, w). \tag{20}
\]

It will be useful in later work to note that the retrospective expectation (12) is given by
\[
\sum_{s=1}^{S} \sum_{d=0}^{1} n_{ds} E \{\mathcal{L}_{is}(\Theta) | D_{is} = d, s\} = \sum_{s=1}^{S} n_s \sum_{d=0}^{1} n_{ds}. \tag{21}
\]
Strictly speaking, retrospective unbiasedness of the estimating equation means that (20) holds for all \(\Theta\) and \(q_s(\cdot)\) 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 \(\ell_{ss}(\cdot, \Theta^*) = q_{ss}(z, x) H_{ss}^2(\cdot) \{1 - H_{ss}(\cdot)\}^{1-d} f(\cdot, \theta_2)\), where \(\Theta^* = (\theta_0^s, \theta_1^s, \theta_2)\), \(q_{ss}(\cdot)\) is the marginal of \((Z, X)\) in the prospective sampling distribution in the \(s\)th stratum, and \(H_{ss}(\cdot)\) is the same as \(H_s(\cdot)\) but with prospective stratum-specific intercepts. Thus, prospective unbiasedness means that for all \(\Theta^*\) and \(q_{ss}(\cdot)\) in an appropriate class,
\[
0 = \sum_{s=1}^{S} \sum_{d=0}^{1} \sum_{j=1}^{J} n_s \int \pi_j(\cdot) \Psi_{js}(\cdot, \Theta^*) \ell_{ss}(\cdot, \Theta^*) d\mu(z, x, w). \tag{22}
\]

Note the similarity between (20) and (22). The equations are formally identical, with the only difference being 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 (20) and (22) trivially leads to retrospective unbiasedness of the estimating equation.

12.3 Sketch of Proof of Main Theorem

Consider the retrospective formulation, where the parameter is \(\Theta\). By a Taylor series expansion, 
\[
n^{1/2} \left( \hat{\Theta} - \Theta \right) \approx - \{T_{n\Theta}(\Theta)^{-1} n^{1/2} T_n(\Theta) \}.
\]
By a calculation similar to (21), \(T_{n\Theta}(\Theta)\) has expectation (13): suppressing the dependence on sample sizes, denote the result by \(T_{\Theta}(\Theta)\).
We next compute \( \text{cov}\{n^{1/2}T_n(\Theta)\} \) (conditioned on all the \( D \)'s, of course). Let the notation \([...]\) indicate a repeat of the preceding term. Using (12),

\[
\text{cov}\{n^{1/2}T_n(\Theta)\} = n^{-1} \sum_{s=1}^{S} \sum_{i=1}^{n_s} E \left( \left[ \mathcal{L}_{is}(\Theta) - E\{\mathcal{L}_{is}(\Theta)\} \right] \left[ ... \right] | D_{is}, s \right)
\]

\[
= n^{-1} \sum_{s=1}^{S} \sum_{i=1}^{n_s} \left( E\{\mathcal{L}_{is}(\Theta) \mathcal{L}_{is}^t(\Theta)\} | D_{is}, s \} - \left[ E\{\mathcal{L}_{is}(\Theta)\} | D_{is}, s \}] \left[ ... \right] \right)\]

\[
= C(\Theta) \left( \sum_{s=1}^{S} \sum_{d=1}^{1} \frac{1}{n} \right) E\{\mathcal{L}_{is}(\Theta) \mathcal{L}_{is}^t(\Theta) | D_{is} = d, s\} \right] \left[ ... \right].
\]

It is easily seen that \( E\{\mathcal{L}_{is}(\Theta) | D_{is} = d, s\} = (n_s/n_d)\kappa_{ds} \), thus showing that

\[
\text{cov}\{n^{1/2}T_n(\Theta)\} = C(\Theta) - \sum_{s=1}^{S} \sum_{d=1}^{1} \left( \frac{n_s^2}{n d} \right) \kappa_{ds} \kappa_{ds}^t.
\]

Now, employing (10),

\[
C(\Theta) = \sum_{s=1}^{S} \sum_{d=1}^{1} \frac{1}{n} E\{\mathcal{L}_{is}(\Theta) \mathcal{L}_{is}^t(\Theta) | D_{is} = d, s\}
\]

\[
= \sum_{s=1}^{S} \sum_{d=1}^{1} \frac{1}{n} \sum_{j=1}^{J} \sum_{j=1}^{J} \int \pi_j(\cdot) \psi_{j, s}(\cdot, \Theta) \psi_{j, s}^t(\cdot, \Theta) \lambda(\cdot) d\mu(z, x, w).
\]

This is identical to (14), as required.

### 12.4 Theory for the Classical Model

We have defined conditional unbiasedness to mean that (18) holds for all \( \Theta^* \). Write \( \Theta = (\theta_0, \theta_1)' \), \( H(x, \Theta) = H\{\theta_0 + R(\theta, x)\} \) and \( H^1(x, \Theta) = H(x, \Theta) \{1 - H(x, \Theta)\} \). Then conditional unbiasedness means that for any \( (x, \Theta) \),

\[
0 = \sum_{d=0}^{1} \psi(d, x, \Theta) H^d(x, \Theta) \{1 - H(x, \Theta)\}^{1-d}.
\]

In our notation, \( \kappa_d = \int \psi(d, x, \Theta) H^d(x, \Theta) \{1 - H(x, \Theta)\}^{1-d} q(x)dx \), so that \( \kappa_0 + \kappa_1 = 0 \) by (25), and hence prospective estimating equations which are conditionally unbiased are also unbiased retrospectively. It also follows that

\[
\mathcal{T}_\Theta(\Theta) = \int \sum_{d=0}^{1} \psi_d(x, \Theta) H^d(x, \Theta) \{1 - H(x, \Theta)\}^{1-d} q(x)dx.
\]

Differentiating the right hand side of (25) with respect to \( \Theta \) and then integrating with respect to \( q(x)dx \), we find that the first column of \( -\mathcal{T}_\Theta(\Theta) \) is

\[
\int \sum_{d=0}^{1} (2d - 1) \psi(d, x, \Theta) H(x, \Theta) \{1 - H(x, \Theta)\} q(x)dx.
\]
It follows then that $\kappa_1$ equals the first column (26) 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 (25).

Based on the Lemma in section 4, we have thus shown that prospectively defined standard errors for slope parameters are retrospectively asymptotically correct for the general class of conditionally unbiased prospective estimating equations.

12.5 Theory for Differential Error

We first briefly sketch an argument showing that the estimating equations of section 6 are retrospectively unbiased. Since there is only a single stratum, we drop the stratum indicators, and as in section 6 write $\kappa_d = \{\kappa_d(\psi) + \kappa_d(\phi)\}^t, \kappa_d^t(\chi)$. We will show that $\kappa_0(\chi) + \kappa_1(\chi) = 0$, the other cases being similar. By definition,

$$
\sum_{d=0}^1 \kappa_d(\chi) = \sum_{d=0}^1 \int \pi(d, z, w) H^d(\cdot) \{1 - H(\cdot)\}^{1-d} f(w | z, x, d, \Theta) q(z, x) \times \left\{ \chi(z, x, w, d, \Theta) - \frac{G(z, x, w, \pi \chi, \Theta)}{G(z, x, w, \pi, \Theta)} \right\} d\mu(z, x, w)
$$

$$
= \int \left\{ \frac{G(\cdot, \pi \chi, \Theta)}{G(\cdot, \pi, \Theta)} - \frac{G(\cdot, \pi \chi, \Theta)}{G(\cdot, \pi, \Theta)} \sum_{d=0}^1 \pi(\cdot) H^d(\cdot) \{1 - H(\cdot)\}^{1-d} f(\cdot, \Theta) \right\} q(\cdot) d\mu(z, x, w).
$$

Since this integrand is zero, this yields the desired result.

We now show that except for the intercept, prospective covariance formulae are asymptotically correct. To do this, we must show that $\kappa_1$ is proportional to the first column of $T_\Theta(\Theta)$. Let $H^{(1)}(\cdot) = H(\cdot) \{1 - H(\cdot)\}$ be the derivative of $H(\cdot)$. Since $G\{\cdot, \xi M(z, x), \Theta\} = M(z, x) G(\cdot, \xi, \Theta)$, direct calculations indicate that $\kappa_1 = \{\kappa_1(\psi) + \kappa_1(\phi)\}^t, \kappa_1^t(\chi)$, where

$$
\kappa_1(\psi) = \int \pi(1, z, w) M(z, x) f(w | z, x, d = 1, \Theta) q(z, x) \times \left[ H^{(1)}(\cdot) - \frac{H(\cdot)G\{\cdot, \pi (1 - H(\cdot), \Theta)\}}{G(\cdot, \pi, \Theta)} \right] d\mu(z, x, w);
$$

$$
\kappa_1(\phi) = \int H(\cdot) f(w | z, x, d = 1, \Theta) q(z, x) \times \left[ 1 - \pi(1, z, w) \phi(1, z, w) - \pi(1, z, w) \frac{G(\cdot, (1 - \pi) \phi, \Theta)}{G(\cdot, \pi, \Theta)} \right] d\mu(z, x, w);
$$

$$
\kappa_1(\chi) = \int \pi(1, z, w) H(\cdot) f(w | z, x, d = 1, \Theta) q(z, x) \times \left[ \chi(w, z, x, d = 1, \Theta) - \frac{G(\cdot, \pi \chi, \Theta)}{G(\cdot, \pi, \Theta)} \right] d\mu(z, x, w).
$$

With $\theta_0$ being the intercept, $(\partial / \partial \theta_0) \Psi_2 = 0$, $(\partial / \partial \theta_0) \psi = M(z, x) H^{(1)}$, $(\partial / \partial \theta_0) H^{d(1 - H)^{1-d}} = (2d - 1) H^{(1)}$, and for any function $\xi(\cdot)$, $(\partial / \partial \theta_0) G(\cdot, \xi e, \Theta) = G(\cdot, \xi e, \Theta) + G(\cdot, (d - H) \xi, \Theta)$. Writing $G(\cdot, \xi, \Theta) = G(\xi)$, by direct but tedious algebra, we find that if $\Psi_1 = (\psi_{ia}^t, \psi_{ib}^t)^t$, then

$$
G(\pi)(\partial / \partial \theta_0) \psi_{ia} = -M(z, x) \left[ G(\pi (d - H)^2) - \frac{G^2(\pi (d - H))}{G(\pi)} \right].
$$
\[ - \left[ G\{1 - \pi)(d - H)\phi\} - \frac{G\{(1 - \pi)\phi\} G\{\pi(d - H)\}}{G(\pi)} \right]. \] (27)

Clearly, \((\partial/\partial \theta_0) \psi_{1a}\) does not depend on \(d\). Remembering that \((\partial/\partial \theta_0) \Psi_2 = 0\), the part of the first column of \(T_0(\Theta)\) corresponding to \(\psi_{1a}\) is

\[
\sum_{d=0}^{1} \int \pi(\cdot) (\partial/\partial \theta_0) \psi_{1a}(\cdot) \cdot H^d(\cdot) \{1 - H(\cdot)\}^{1-d} f(\cdot|\cdot, \Theta) q(\cdot) d\mu(z, x, w)
= \int \mathcal{G}(\cdot, \pi, \Theta) (\partial/\partial \theta_0) \psi_{1a}(\cdot) q(\cdot) d\mu(z, x, w). \] (28)

We now substitute the right side of (27) into (28), noting that it factors naturally into components depending on \(\phi\) and \(\psi\), the latter through \(M\). We thus rewrite (28) as

\[
- \int \{G_1(\psi) + G_2(\phi)\} q(\cdot) d\mu(z, x, w). \] (29)

Direct but tedious algebra shows that the integrands corresponding to \(\psi\) and \(\phi\) in (29) exactly equal the integrands in \(\kappa_1(\psi)\) and \(\kappa_1(\phi)\).

Similarly, the part of the first column of \(T_0(\Theta)\) corresponding to \(\psi_{1b}\) is

\[
- \int \left[ \mathcal{G}(\cdot, \pi X(d - H), \Theta) - \frac{\mathcal{G}(\cdot, \pi X, \Theta) \mathcal{G}(\cdot, \pi(d - H), \Theta)}{\mathcal{G}(\cdot, \pi, \Theta)} \right] q(\cdot) d\mu(z, x, w).
\]

It can be shown that the integrand of this expression equals the integrand of \(\kappa_1(\chi)\).

We have thus shown that \(\kappa_1\) is proportional to the first column of \(T_0(\Theta)\), and hence that prospective covariance formulae may be used.

12.6 Theory for Nondifferential Measurement Error

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

\[
(n/n_d) q(z, w) H_R^d(\cdot) \{1 - H_R(\cdot)\}^{1-d} f_{X|Z,W,D}(z|z, w, d, \theta_2);
H_R(\cdot) = H\{\theta_0 + \theta_1 z + R(z, w, \Theta)\}.
\]

In our theory, \(H_R(\cdot)\) replaces \(H(\cdot)\), and \(f_{X|Z,W,D}(\cdot)\) replaces \(f(\cdot)\).

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

\[
\Psi_1(\cdot) = \left( \mathcal{M}(\cdot) \{D - H_R(\cdot)\}, S^t_{\theta_1}, S^t_{\theta_2} \right)^t;
\Psi_2(\cdot) = \left( \mathcal{M}(\cdot) \{D - H_R(\cdot)\}, 0^t, 0^t \right)^t,
\]

where \(S_{\theta_2} = (\partial/\partial \theta_2) \log \left\{ f_{X|Z,W,D}(X|Z,W,D,\Theta) \right\}\), \(S_{\theta_1}\) is defined similarly. In addition, \(\mathcal{M}(Z,W) = (1, Z^t, U_{\theta_{1z}}, U_{\theta_{2z}})^t\), where \(U_{\theta_2} = (\partial/\partial \theta_2) R(Z,W,\Theta)\) and similarly for \(U_{\theta_{1z}}\).
It is easy to show that these prospective estimating equations are retrospectively unbiased. With the redefinition of \( \ell(\cdot) \), the first column of \( \mathcal{T}_0(\Theta) \) is

\[
\mathcal{T}_{01}(\Theta) = - \left( \sum_{d=0}^{1} \int \mathcal{M}(z, w) H_R^{(1)}(\cdot) \ell(\cdot, \Theta) d\mu(z, x, w), 0^t, 0^t \right)^t
\]

\[
= - \left( \sum_{d=0}^{1} \int \mathcal{M}(z, w) H_R^{(1)}(\cdot) H_R^{(d)}(\cdot) \left\{ 1 - H_R(\cdot) \right\}^{1-d} \times q(z, w) f_{X|Z,W,D}(z|z, w, d, \Theta) d\mu(z, x, w), 0^t, 0^t \right)^t
\]

\[
= - \left( \int \mathcal{M}(z, w) H_R^{(1)}(\cdot) q(z, w) d\mu(z, x, w), 0^t, 0^t \right)^t,
\] (30)

the last step following because the only term depending on \( x \) is \( f_{X|Z,W,D}(z|z, w, d, \Theta) \), which is a density and hence integrates to one. That (30) equals \( -\kappa_1 \) is immediate.

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

### 12.7 Theory for Corrections for Attenuation

Let the mean of \( W_1 \) among the controls be \( \mu_w \) and the mean of \( W_1 - W_2 \) among the controls be \( \mu_e = 0 \). For technical reasons having to do with the fact that \( \sigma_u^2 \) is being estimated by the sample variance \( \hat{\sigma}_u^2 \), we must include an estimating equation for \( \mu_e \) even though it is known; the estimating equation has no effect on the standard error estimates. The estimating equations for this algorithm are, with \( \theta_2 = (\mu_w, \mu_e, \sigma_w^2, \sigma_u^2)^t \),

\[
\Psi_{11} = \begin{bmatrix}
1 & 0 \\
0 & g_1(W_1, \sigma_w^2, \sigma_u^2, \mu_w) \left( W_1 - \mu_w \right) I(D = 0) \\
0 & \left\{ (W_1 - \mu_w)^2 - \sigma_w^2 \right\} I(D = 0) \\
0 & 0
\end{bmatrix}
\]

and

\[
\Psi_{12} = \begin{bmatrix}
0 & 1 \\
0 & g_2(W, \sigma_w^2, \sigma_u^2, \mu_w) \left( W_1 - \mu_w \right) I(D = 0) \\
0 & \left\{ (W_1 - \mu_w)^2 - \sigma_w^2 \right\} I(D = 0) \\
0 & \left\{ (W_1 - W_2) / 2^{1/2} - \mu_e \right\} \\
0 & \left\{ (W_1 - W_2) / 2^{1/2} - \mu_e \right\}^2 - \sigma_u^2
\end{bmatrix}
\]

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 \( X \), since \( \kappa_{1a} \) is based on the case \( d = 1 \), it follows trivially that

\[
\kappa_{1s} = \int \{ a, b, g_s(\cdot), 0, 0, 0 \}^t H_R^{(1)}(\cdot) q(\cdot) f(w|x) d\mu(z, w),
\]

25
where \((a, b) = (1, 0)\) and \((0, 1)\) for \(s = 1, 2\), respectively. It is also easily verified that the first column of \(\mathcal{T}_\Theta(\Theta)\), which is based on the expectations of the derivative of \(\Psi_{11}\) with respect to \(\theta_{01}\), equals \(-\kappa_{11}\), while the second column of \(\mathcal{T}_\Theta(\Theta)\) is \(-\kappa_{12}\). Hence, subject to the levels of approximation described, the conclusion is that prospective covariance formulae may be used for the estimate of \(\theta_1\).

The way we estimated standard errors is by using a prospective formulation. The covariances of the terms \(\Psi_{11}\) and \(\Psi_{12}\) were originally computed using the “model-free” method, with the usual exception that in the upper \(3 \times 3\) 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 \(\Psi_{11}\) and \(\Psi_{12}\), we again started with the model-free method, and we again modified the upper \(3 \times 3\) matrix by substituting information contributions. For the other terms in the first three rows of \(\Psi_{11}\) and \(\Psi_{12}\), 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

\[
\left\{ \frac{\partial}{\partial \sigma^2} g_1(W_1, \sigma^2_w, \sigma^2_u, \mu_w) \right\}^T \left[ D - H \left\{ \theta_{01} + \theta_1 g_1(W_1, \sigma^2_w, \sigma^2_u, \mu_w) \right\} \right]
\]

have mean zero. Generalization to vector predictors is immediate.

### 12.8 Theory for Partial Questionnaires

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

\[
G_4 = Hq; \quad G_3 = \int Hq d\mu(x_1); \quad G_2 = \int Hq d\mu(x_2); \quad G_1 = \int Hq d\mu(x_1, x_2);
\]

\[
A_4 = q; \quad A_3 = \int q d\mu(x_1); \quad A_2 = \int q d\mu(x_2); \quad A_1 = \int q d\mu(x_1, x_2);
\]

\[
C_j(\cdot) = \{ \partial/\partial(\theta_{01}, \theta_{11}, \theta_{12}, \theta_{13}) \} G_j(\cdot); \quad L_j(\cdot) = (\partial/\partial \theta_2) G_j(\cdot);
\]

\[
M_j(\cdot) = (\partial/\partial \theta_2) A_j(\cdot); \quad M(\cdot) = (1, x_1, x_2, z)^t.
\]

The prospective estimating equations are in our general form with

\[
\Psi_{j1} = \left\{ \begin{array}{c} C_j(\cdot) \{ DA_j(\cdot) - G_j(\cdot) \} / \{ G_j(\cdot) \{ A_j(\cdot) - G_j(\cdot) \} \} \\ DL_j(\cdot) / G_j(\cdot) + (1 - D) \{ M_j(\cdot) - L_j(\cdot) \} / \{ A_j(\cdot) - G_j(\cdot) \} \end{array} \right\}.
\]

The estimating equation is retrospectively unbiased, and by definition we write

\[
\kappa_{11} = (\kappa_{11a}, \kappa_{11b})^T = \sum_{j=1}^4 \int \pi_j(z, 1) \Psi_{j1}(1, z, x_1, x_2, \Theta) H(\cdot) q(\cdot) d\mu(x_1, x_2, z).
\]

Define \(d\mu_4(\cdot) = d\mu(x_1, x_2, z); d\mu_3(\cdot) = d\mu(x_2, z); d\mu_2(\cdot) = d\mu(x_1, z); d\mu_1(\cdot) = d\mu(z).\) Recall that \((\partial/\partial v) H(v) = H^{(1)}(v) = H(v) \{1 - H(v)\}\) and define \(R(\cdot) = (\partial/\partial \theta_2) q(\cdot).\) Then

\[
\kappa_{11a} = \sum_{j=1}^4 \int \pi_j(z, 1) \{ C_j(\cdot) / G_j(\cdot) \} H(\cdot) q(\cdot) d\mu(x_1, x_2, z).
\]
Note, however, that $C_j(\cdot)$ and $G_j(\cdot)$ depend only on $z, (x_1, z), (x_2, z)$ and $(x_1, x_2, z)$ for $j = 1, 2, 3, 4$, respectively, so that

$$\kappa_{11a} = \sum_{j=1}^{4} \int \pi_j(z, 1) C_j(\cdot) d\mu_j(\cdot) = \sum_{j=1}^{4} \int \pi_j(z, 1) C_4(\cdot) d\mu(x_1, x_2, z)$$

$$= \int C_4(\cdot) d\mu(x_1, x_2, z) = \int M(\cdot) H^{(1)}(\cdot) q(\cdot) d\mu(x_1, x_2, z).$$

Similarly,

$$\kappa_{11b} = \sum_{j=1}^{4} \int \pi_j(z, 1) \left\{ L_j(\cdot)/G_j(\cdot) \right\} H(\cdot) q(\cdot) d\mu(x_1, x_2, z)$$

$$= \int L_4(\cdot) d\mu(x_1, x_2, z) = \int H(\cdot) R(\cdot) d\mu(x_1, x_2, z).$$

Because $T_0(\Theta)$ and $C(\Theta)$ of (13)–(14) are the same as in the prospective case, and the estimating equations are obtained from maximum likelihood, it follows that $T_0(\Theta) = -C(\Theta)$; this may be verified directly by algebra. It is easier to compute $C(\Theta)$, which is given by

$$C(\Theta) = \sum_{j=1}^{4} \int \pi_j(z, 1) \left( \frac{C_j(\cdot)}{L_j(\cdot)} \right) \left( \frac{C_j(\cdot)}{L_j(\cdot)} \right)^t \{ G_j(\cdot) \}^{-1} d\mu_j(\cdot)$$

$$+ \sum_{j=1}^{4} \int \pi_j(z, 1) \left( \frac{-C_j(\cdot)}{M_j(\cdot) - L_j(\cdot)} \right) \left( \frac{-C_j(\cdot)}{M_j(\cdot) - L_j(\cdot)} \right)^t \{ A_j(\cdot) - G_j(\cdot) \}^{-1} d\mu_j(\cdot).$$

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.

12.9 Theory for Two-Stage Studies

Let $n_d$ be the number of observations with $D = d$, let $n_{md}$ be the random number of observations with $(D = d, Z = m)$ and let $n_{md}^*$ be the fixed number of observations in the second stage within each $(D, Z)$ category. Define $\theta_{2,md} = \text{pr}(Z = m|D = d)$. Note that $(n_{1d}, \cdots, n_{Md})$ is a multinomial random variable with probabilities $(\theta_{2,1d}, \cdots, \theta_{2,Md})$.

Because there is only a single stratum, we will drop the stratum assignment indicators. In (11), $j = 1$ refers to observations selected into the second stage sample, while $j = 2$ denotes those which are not so selected. Define $\psi_{21}(\cdot, \Theta) = 0$ and

$$\psi_{11}(\cdot, \Theta) = (1, X)^t \left[ d - H \left\{ \theta_0 + \theta_1 X + \log (n_0/n_1) + \log (n_{Z1}/n_{Z0}) + \log (\theta_{2,0}/\theta_{2,Z1}) \right\} \right]. \quad (31)$$

Let $\psi_{12} = \psi_{22}(\cdot, \Theta)$ be the vector of size $2M$ whose $(dM+m)^{th}$ element equals $I(D = d) \{ I(Z = m) - \theta_{2,md} \}$.

Let $\Psi_j = (\psi_{j1}, \psi_{j2})^t$. Denote the logistic argument in (31) by $H_* (Z, X, \Theta)$ and write $H_*^{(1)} = H_*(1 - H_*)$.

At the end of this section, the estimating equation is shown to be unbiased, the particular method being to condition on all $n_{md} \geq n_{md}^*$, or equivalently on all the $\delta$'s. In addition, if $(\tilde{D}, \tilde{Z})$ denotes the collection of all $(D, Z)$'s, we later show that

$$0 = E \left\{ n^{-1} \sum_{i=1}^{n} \delta_{i1} \psi_{11}(\cdot, \Theta) | \tilde{D}, \tilde{Z} \right\}. \quad (32)$$
Next we show that the estimating equations for \((\theta_0, \theta_1)\) are uncorrelated with those for \((\theta_2, \theta_3, \cdots, \theta_{2,M_1})\), 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

\[
n^{-1} E \left( \sum_{i=1}^{n} \delta_{11} \psi_{11} (\cdot, \Theta) \left\{ \sum_{i=1}^{n} \delta_{11} \psi_{12} (\cdot, \Theta) + n^{-1} \sum_{i=1}^{n} \delta_{12} \psi_{22} (\cdot, \Theta) \right\} | \tilde{D}, \tilde{Z} \right).
\]

The terms associated with \(\psi_{12} = \psi_{22}\) depend only on \((\tilde{D}, \tilde{Z})\), so that if we condition on this term and apply (32), we have the desired result.

Breslow & Cain do not use our estimating equation approach, but their results are equivalent to ours, except that they work with the parameterization \(\xi_{md} = \log(\theta_{2,md})\). In our notation, we have shown that

\[
T_\Theta(\Theta) = \begin{bmatrix} T_{\Theta 11} & T_{\Theta 12} \\ 0 & T_{\Theta 22} \end{bmatrix}; \quad A(\Theta) = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}.
\]

Except for the change in parameterization and the fact that our asymptotics are based on the total sample size \(n\) rather than the total \(n_* = \sum_{m,d} n_{md}\), a notational difference of no effect on the final results, their term \(H\) corresponds to our \(T_{\Theta 11}\), their term \(A\) to our \(T_{\Theta 12}\), their term \(B\) to our \(A_{22}\) and, as we show at the end of this section, their term \(G\) to our \(A_{11}\).

We now turn to filling in the main technical steps. We first show (32). The notation is that \(H_*(\cdot, \Theta)\) refers to the logistic argument in (31). Since the conditional density or probability mass function \(f(x|z, d) = f(x, z|d)/\theta_{2,ad}\), then

\[
E \left[ \sum_{i=1}^{n} \delta_{11} (1, X_i^t)^t \{ D_i - H_*(\cdot, \Theta) \} | \tilde{D}, \tilde{Z}, \delta \right]
\]

\[
= \sum_{m=1}^{M} \sum_{d=0}^{1} n_{md} \text{E} \left[ (1, X_i^t)^t \{ d - H_*(\cdot, \Theta) \} | \tilde{D}, \tilde{Z} \right]
\]

\[
= \sum_{m=1}^{M} \sum_{d=0}^{1} n_{md} \text{E} \left[ (1, X_i^t)^t \{ d - H_*(\cdot, \Theta) \} | \tilde{D}, \tilde{Z} \right]
\]

\[
= \sum_{m=1}^{M} \int (1, x^t)^t q(x, m) \frac{1}{n_d} \frac{n_{md}}{\theta_{2,md}} \{ d - H_*(\cdot) \} H^d(\cdot) \{ 1 - H(\cdot) \} (1-d) d\mu(x).
\]

(33)

However, the term in square brackets in (33) equals zero, proving (32).

Next we prove that the estimating equation is unbiased. The part corresponding to \(\psi_{11}(\cdot, \Theta)\) is unbiased by (32). For the other part, for specificity consider the estimating equation corresponding to \(\theta_{2,m_1}\). This has expectation

\[
(n_1/n) \int \{ I(z = m) - \theta_{2,m_1} \} (n/n_1) H(\cdot) q(x, z) d\mu(x, z).
\]

Since \((n/n_1) H(\cdot) q(x, z)\) is the distribution of \((X, Z)\) given \(D = 1\), the estimating equation has expectation \((n_1/n) \{ \text{pr}(Z = m | D = 1) - \theta_{2,m_1} \} = 0\).

Now, we show that our \(A_{11}\) is the same as Breslow & Cain's matrix \(G\) except for the replacement in our calculations of \(n^{-1}\) by their \(n_*^{-1}\), where \(n_* = \sum_{z,d} n_{zd}\). Define \(\xi(\cdot, \Theta) = E \{ \psi_{11}(\cdot, \Theta) | \tilde{D}, \tilde{Z} \} \).
We have that
\[
A_{11}(\Theta) = n^{-1} \text{cov} \left\{ \sum_{i=1}^{n} \delta_{i1} \psi_{11}(\cdot, \Theta) | \bar{D} \right\}
\]
\[
= n^{-1} E \left[ \text{cov} \left\{ \sum_{i=1}^{n} \delta_{i1} \psi_{11}(\cdot, \Theta) | \bar{D}, \bar{Z}, \delta \right\} | \bar{D} \right] + n^{-1} \text{cov} \left[ E \left\{ \sum_{i=1}^{n} \delta_{i1} \psi_{11}(\cdot, \Theta) | \bar{D}, \bar{Z}, \delta \right\} | \bar{D} \right]
\]
\[
= n^{-1} E \left[ \text{cov} \left\{ \sum_{i=1}^{n} \delta_{i1} \psi_{11}(\cdot, \Theta) | \bar{D}, \bar{Z}, \delta \right\} | \bar{D} \right],
\]
the last step following from (33). Thus,
\[
A_{11}(\Theta) = n_\ast^{-1} E \left( E \left[ \sum_{i=1}^{n} \delta_{i1} \left\{ \psi_{11}(\cdot, \Theta) \psi_{11}(\cdot, \Theta) - \xi(\cdot, \Theta) \xi(\cdot, \Theta) \right\} | \bar{D}, \bar{Z}, \delta \right] | \bar{D} \right)
\]
\[
= \sum_{e=1}^{M} \sum_{d=0}^{1} (n_\ast^{d}/n_\ast) E \left\{ \psi_{11}(\cdot, \Theta) \psi_{11}(\cdot, \Theta) - \xi(\cdot, \Theta) \xi(\cdot, \Theta) | D = d, Z = z \right\}.
\]
This last term is Breslow & Cain’s matrix \( G \).