

Covariate Measurement Error Adjustment for Matched Case-Control Studies

Lisa M. McShane,^{1,*} Douglas N. Midthune,² Joanne F. Dorgan,³
Laurence S. Freedman,⁴ and Raymond J. Carroll⁵

¹National Cancer Institute, Biometric Research Branch, DCTD,
Executive Plaza North, Room 739, 6130 Executive Boulevard, MSC 7434,
Bethesda, Maryland 20892-7434, U.S.A.

²National Cancer Institute, Biometry Research Group, DCP,
Executive Plaza North, Room 344, 6130 Executive Boulevard, Bethesda, Maryland 20892, U.S.A.

³Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, Pennsylvania 19111, U.S.A.

⁴Department of Mathematics, Statistics, and Computer Science, Bar Ilan University, Ramat Gan 52900, Israel

⁵Department of Statistics, Texas A&M University, College Station, Texas 77843-3143, U.S.A.

*email: lm5h@nih.gov

SUMMARY. We propose a conditional scores procedure for obtaining bias-corrected estimates of log odds ratios from matched case-control data in which one or more covariates are subject to measurement error. The approach involves conditioning on sufficient statistics for the unobservable true covariates that are treated as fixed unknown parameters. For the case of Gaussian nondifferential measurement error, we derive a set of unbiased score equations that can then be solved to estimate the log odds ratio parameters of interest. The procedure successfully removes the bias in naive estimates, and standard error estimates are obtained by resampling methods. We present an example of the procedure applied to data from a matched case-control study of prostate cancer and serum hormone levels, and we compare its performance to that of regression calibration procedures.

KEY WORDS: Case-control study; Conditional logistic regression; Conditional scores; Hormones; Matched design; Measurement error; Prostate cancer.

1. Introduction

There has been a proliferation of biorepositories holding serum or tissue specimens collected from subjects in large clinical trials or prospectively followed cohorts. Collected prediagnosis, these specimens can be used to examine relationships between risk of disease and serum and tissue biomarkers measured by laboratory assays. The nested case-control design, which involves matching on characteristics that might otherwise confound exposure-disease relationships, is frequently used for such studies. Typically, one has only a single measurement of the biomarker per individual and it may be subject to measurement error arising from multiple sources. We envision that each subject has a true underlying average measure for the biomarker of interest. The actual level on any occasion may vary from this average for numerous reasons. For example, for biomarkers measured in serum, biological variation related to inherent patterns of secretion (e.g., diurnal rhythms) or changes in personal characteristics (e.g., diet) that are unmeasured or unknown to affect the biomarker of interest cause fluctuations in levels. Differences in specimen collection or handling may also cause fluctuations. We refer to the combined effects of random biological variation and

specimen handling on biomarker levels as occasion-within-person variability. There is also laboratory assay variability, which may be subdivided into between-batch (i.e., assay) and within-batch variability. We consider the contributions of all of these sources of variability as measurement error with regard to an individual's true biomarker level.

Measurement error in an explanatory exposure variable may result in attenuation of relative risk estimates and reduced power for detecting exposure-disease relationships. For unmatched studies, a variety of measurement error correction methods have been proposed for logistic risk models (Rosner, Willett, and Spiegelman, 1989; Rosner, Spiegelman, and Willett, 1990, 1992; Carroll, Ruppert, and Stefanski, 1995, and references therein), but the matched design has received far less attention. Armstrong, Whittemore, and Howe (1989) propose a measurement error correction method assuming a normal discriminant analysis model. Their method assumes multivariate Gaussian covariates, but in that setting, it has the flexibility to handle differential measurement error. Forbes and Santner (1995) develop a correction method using a retrospective likelihood with a binary exposure variable. Their method assumes that the binary exposure variable is mea-

sured without error and addresses situations in which continuous confounders may be measured with error. Prentice (1982) proposes a method for parameter estimation in Cox's failure time regression model when the covariates are measured with error. Noting the similarity between Cox's partial likelihood and the conditional logistic regression likelihood used for matched analyses, Prentice's method is applicable here. But its implementation requires knowledge of the conditional distribution of the true covariates given their observed error-prone measurements or, at a minimum, sufficient information to compute the conditional expectation of the exponential terms in the likelihood.

The conditional scores method we propose for matched studies is based on the prospective likelihood, and it allows for very general covariate distributions. In this approach, unobservable true covariates are treated as fixed unknown parameters. One condition on a sufficient statistic to remove them from the likelihood and produce a set of unbiased score equations to solve for log odds ratio parameter estimates.

In Section 2, we review the matched case-control study design and likelihood. A measurement error model appropriate for continuous biomarker variables measured by laboratory assay is presented in Section 3. In Section 4, the conditional scores measurement error adjustment procedure is developed under an assumption of Gaussian nondifferential measurement error. Regression calibration approaches are described in Section 5. In Section 6, we apply the procedures to example data from a study examining the relationship between serum hormone levels and risk of prostate cancer. Simulation studies are presented in Section 7. A discussion follows in Section 8.

2. Study Design

The study design considered here is a 1: M matched case-control study with K strata (matched sets), although the results can be extended to more general matching. The disease outcome is a binary variable with logit of disease probability modeled by $g_k(\mathbf{r}) = a_k + \beta' \mathbf{r}$, where a_k denotes the contribution of all terms constant within stratum k , \mathbf{r} is a vector of p covariates, and $\beta = (\beta_1, \beta_2, \dots, \beta_p)'$ is a coefficient vector.

The usual method of analysis is conditional logistic regression. Associated with each subject there is a p -dimensional covariate vector, denoted by \mathbf{r}_{k1} for the case in stratum k and by $\mathbf{r}_{k2}, \mathbf{r}_{k3}, \dots, \mathbf{r}_{k,M+1}$ for the M controls. When all covariates are measured without error, the conditional prospective logistic likelihood for 1: M matching (cf., Hosmer and Lemeshow, 1989) is

$$l(\beta) = \prod_{k=1}^K \left\{ 1 + \sum_{j=2}^{M+1} e^{(\mathbf{r}_{kj} - \mathbf{r}_{k1})' \beta} \right\}^{-1} \quad (1)$$

For 1:1 matching, this reduces to the likelihood function for unconditional logistic regression with no intercept term, with individual covariate vectors replaced by control minus case differences and with all responses set as zero. When there is measurement error in covariates, tests and estimates of β derived from usual maximum likelihood techniques applied to (1) are biased (Armstrong et al., 1989). Corrections for that bias in the general setting of 1: M matching are the subject of this article.

3. Measurement Error Model

Suppose that the first p_1 components of each covariate vector, \mathbf{r}_{kj} , are measured with error. We denote the true (unobservable, error-free) components by $\mathbf{x}_{kj} = (x_{kj}^{(1)}, x_{kj}^{(2)}, \dots, x_{kj}^{(p_1)})'$ and the error-prone, observable version of \mathbf{x}_{kj} by \mathbf{w}_{kj} . The remaining $p_2 = p - p_1$ components of \mathbf{r}_{kj} , denoted by $\mathbf{z}_{kj} = (z_{kj}^{(1)}, z_{kj}^{(2)}, \dots, z_{kj}^{(p_2)})'$, are observed without error. Let $u_{kj}^{(i)}$ represent additive measurement error on the variable $x_{kj}^{(i)}$ such that $w_{kj}^{(i)} = x_{kj}^{(i)} + u_{kj}^{(i)}$, $i = 1, 2, \dots, p_1$; $j = 1, 2, \dots, M+1$; $k = 1, 2, \dots, K$. We require that, for each i , $\{u_{kj}^{(i)}\}$, over all k and j have constant mean and variance and be independent of $\{x_{kj}^{(i)}\}$ and also of $\{z_{kj}^{(i)}\}$ and logits $\{g_k(\mathbf{r})\}$; therefore, the $\{u_{kj}^{(i)}\}$ satisfy the conditions of nondifferential measurement error.

We model the measurement error as

$$u_{kj}^{(i)} = c^{(i)} + O_{kj}^{(i)} + B_k^{(i)} + \epsilon_{kj}^{(i)} \quad (2)$$

for $j = 1, 2, \dots, M+1$; $k = 1, 2, \dots, K$; $i = 1, 2, \dots, p_1$, where $c^{(i)}$ is a constant depending only on i , $\{O_{kj}^{(i)}\}$ are random occasion-within-person effects, $\{B_k^{(i)}\}$ are random laboratory batch effects, and $\{\epsilon_{kj}^{(i)}\}$ are random within-batch error effects. To eliminate assay batch effects, it is standard practice to run all samples from the case and controls from a matched set together in a batch. The notation and description here require this batching design. The batch effect corresponding to the batch containing the k th stratum samples for measuring covariate i will be denoted by $B_k^{(i)}$. (The number of distinct batch effects is typically fewer than the number of matched sets.) All random effects are assumed to have mean zero. The occasion effects, batch effects, and within-batch errors are independent of each other. The $\{\epsilon_{kj}^{(i)}\}$ are independent, with variance depending only on i , denoted by $\sigma_E^2(i)$. The effects $O_{kj}^{(i)}$ and $O_{kj}^{(i')}$ may be correlated, with covariance $\sigma_O(i, i') = \text{cov}(O_{kj}^{(i)}, O_{kj}^{(i')})$. For example, blood levels of two or more hormones may be correlated because the hormones share metabolic pathways or are controlled by complex, tightly regulated processes such as feedback loops in which a fluctuation in the level of one hormone signals changes in the levels of other hormones. We assume that different assays are used to measure the different types of markers (covariates) so batch effects are independent between markers. We denote the variance of the (distinct) batch effects for covariate i by $\sigma_B^2(i)$. The quantities $\sigma_O(i, i')$, $\sigma_B^2(i)$, and $\sigma_E^2(i)$ can be estimated from either an internal or appropriate external variability study.

Since likelihood (1) is a function of the control-case covariate differences, we have particular interest in the structure of the measurement error on these covariate difference vectors. For each stratum, we define $p_1 M \times 1$ and $p_2 M \times 1$ vectors

$$\mathbf{d}_{kx} = (\mathbf{x}'_{k2} - \mathbf{x}'_{k1}, \mathbf{x}'_{k3} - \mathbf{x}'_{k1}, \dots, \mathbf{x}'_{k,M+1} - \mathbf{x}'_{k1})'$$

and

$$\mathbf{d}_{kz} = (\mathbf{z}'_{k2} - \mathbf{z}'_{k1}, \mathbf{z}'_{k3} - \mathbf{z}'_{k1}, \dots, \mathbf{z}'_{k,M+1} - \mathbf{z}'_{k1})'$$

The error-prone version of \mathbf{d}_{kx} is \mathbf{d}_{kw} , and it is defined analogously. \mathbf{d}_{ku} satisfies $\mathbf{d}_{kw} = \mathbf{d}_{kx} + \mathbf{d}_{ku}$ and has variance $\Sigma_{\mathbf{d}_u, \mathbf{d}_u}$

with elements

$$\begin{aligned} & \text{cov} \left(u_{kj}^{(i)}, u_{kj'}^{(i')} \right) - \text{cov} \left(u_{kj}^{(i)}, u_{k1}^{(i')} \right) - \text{cov} \left(u_{k1}^{(i)}, u_{kj'}^{(i')} \right) \\ & + \text{cov} \left(u_{k1}^{(i)}, u_{k1}^{(i')} \right), \end{aligned}$$

where

$$\begin{aligned} & \text{cov} \left(u_{kj}^{(i)}, u_{kj'}^{(i')} \right) \\ & = \begin{cases} \text{var} \left(B_k^{(i)} \right) & \text{if } j \neq j' \text{ and } i = i' \\ \text{var} \left(O_{kj}^{(i)} \right) + \text{var} \left(B_k^{(i)} \right) \\ \quad + \text{var} \left(\epsilon_{kj}^{(i)} \right) & \text{if } j = j' \text{ and } i = i' \\ \text{cov} \left(O_{kj}^{(i)}, O_{kj'}^{(i')} \right) & \text{if } j = j' \text{ and } i \neq i' \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

from model (2). Then Σ_{d_u, d_u} is a block matrix of the form

$$\begin{bmatrix} \Sigma_{d_u, 2, 2} & \Sigma_{d_u, 2, 3} & \cdots & \Sigma_{d_u, 2, M+1} \\ \Sigma_{d_u, 3, 2} & \Sigma_{d_u, 3, 3} & \cdots & \Sigma_{d_u, 3, M+1} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{d_u, M+1, 2} & \cdots & \cdots & \Sigma_{d_u, M+1, M+1} \end{bmatrix},$$

in which each element, $\Sigma_{d_u, j, j'}$, is a $p_1 \times p_1$ matrix equal to $2\mathbf{V}$ if $j = j'$ and equal to \mathbf{V} if $j \neq j'$, with the elements of the $p_1 \times p_1$ matrix \mathbf{V} given by

$$v_{ii'} = \begin{cases} \sigma_{O(i, i)} + \sigma_B^2(i) & \text{if } i = i'; i = 1, 2, \dots, p_1 \\ \sigma_{O(i, i')} & \text{if } i \neq i'; i, i' = 1, 2, \dots, p_1. \end{cases}$$

Observe that Σ_{d_u, d_u} does not depend on the batch-to-batch variances $\{\sigma_B^2(i)\}$ due to assaying samples from the same matched set together in a batch. Not matching on batch will result in loss of the simple structure of Σ_{d_u, d_u} , and measurement error on the differences may become correlated across matched sets, making derivation of the conditional scores estimator extremely difficult. (See Appendix A for additional details.)

4. Conditional Scores Measurement Error Adjustment Method

The bias-corrected estimator we derive is an example of a sufficiency estimator as described by Stefanski and Carroll (1987). Unobservable \mathbf{x} variables are treated as unknown parameters. Conditioning on a sufficient statistic removes them from the likelihood, and unbiased score functions are obtained. As stated previously, measurement error is assumed nondifferential with constant mean and variance. In addition, for this derivation, we assume that the measurement error is Gaussian.

Let $\mathbf{Y}_k = (Y_{k1}, Y_{k2}, \dots, Y_{k, M+1})$ denote the vector of binary response variables associated with the $M+1$ subjects in the k th matched set. Let $\beta = (\beta'_x, \beta'_z)'$ represent the partitioning of β associated with \mathbf{x} and \mathbf{z} . A naive method of estimating β would be to apply maximum likelihood methods to

$$l(\beta) = \prod_{k=1}^K \left\{ 1 + \sum_{j=2}^{M+1} e^{(\mathbf{w}_{kj} - \mathbf{w}_{k1})' \beta_x + (\mathbf{z}_{kj} - \mathbf{z}_{k1})' \beta_z} \right\}^{-1} \quad (3)$$

after replacing all \mathbf{x} 's by \mathbf{w} 's, but this does not lead to consistent estimates. More generally, we write (3) as

$$\begin{aligned} & \Pr \left[\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_K \mid \{\mathbf{x}_k, \mathbf{z}_k, (T_k = 1)\}_{k=1}^K \right] \\ & = \prod_{k=1}^K \frac{\exp \left\{ \sum_{j=1}^{M+1} Y_{kj} (\mathbf{x}'_{kj} \beta_x + \mathbf{z}'_{kj} \beta_z) \right\}}{\sum_{j=1}^{M+1} \exp (\mathbf{x}'_{kj} \beta_x + \mathbf{z}'_{kj} \beta_z)}, \quad (4) \end{aligned}$$

where $\mathbf{x}_k = (\mathbf{x}'_{k1}, \mathbf{x}'_{k2}, \dots, \mathbf{x}'_{k, M+1})'$, $\mathbf{z}_k = (\mathbf{z}'_{k1}, \mathbf{z}'_{k2}, \dots, \mathbf{z}'_{k, M+1})'$, and $T_k = \sum_{j=1}^{M+1} Y_{kj}$. In Appendix A, we show that $\Delta_k = \mathbf{d}_{kw} + \Sigma_{d_u, d_u} \mathbf{B}_{kx}$ is sufficient for \mathbf{d}_{kx} , where β_x is treated as though it were known and $\mathbf{B}_{kx} = (Y_{k2} \beta'_x, Y_{k3} \beta'_x, \dots, Y_{k, M+1} \beta'_x)'$. Then the full conditional likelihood reduces to

$$\begin{aligned} & \Pr \left[\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_K \mid (\Delta_k, \mathbf{d}_{kx}, \mathbf{d}_{kz}, T_k = 1)_{k=1}^K \right] \\ & = \prod_{k=1}^K \left\{ 1 + \sum_{j=2}^{M+1} \exp (\gamma'_{kj} \beta_x + \mathbf{d}'_{kz} \beta_z) \right\}^{-1}, \quad (5) \end{aligned}$$

where $\gamma_{kj} = \mathbf{d}_{kjw} - (1/2) \Sigma_{d_u, 2, 2} \beta_x$, $\mathbf{d}_{kjw} = \mathbf{w}_{kj} - \mathbf{w}_{k1}$, and \mathbf{d}_{kz} are defined analogously. The β_x and β_z that maximize (5) are the solutions to unbiased score equations when the $\{\Delta_k\}$ are held fixed. Under regularity conditions, there exists a consistent solution to these unbiased estimating equations. In Appendix B, we present numerical solution methods that can be performed utilizing standard conditional logistic regression software.

To obtain standard error estimates, one can use any number of methods, including the jackknife. Let $\hat{\beta}_{(k)}$ denote the measurement error corrected estimate of β computed from the full data set minus the k th matched set. The jackknife covariance estimate is

$$\widehat{\text{cov}}_{\text{jack}}(\hat{\beta}) = (K-1) \sum_{k=1}^K (\hat{\beta}_{(k)} - \hat{\beta}_{(\cdot)}) (\hat{\beta}_{(k)} - \hat{\beta}_{(\cdot)})' / K, \quad (6)$$

where $\hat{\beta}_{(\cdot)} = \sum_{k=1}^K \hat{\beta}_{(k)} / K$, and $\hat{\beta}$ is the measurement error corrected estimate using the full data. If measurement error variances are estimated, we show in Section 6 how to obtain variance estimates that incorporate variability due to both the uncertainty in the measurement error variance estimates and the sampling of matched sets.

5. Regression Calibration Measurement Error Adjustment Method

Regression calibration is a general method of measurement error correction useful in a variety of settings, so it is of interest to compare its performance with the conditional scores method. Regression calibration and its broad applicability are described in Carroll et al. (1995). Rosner et al. (1989, 1990, 1992) described an adjustment method for (unconditional) logistic regression that is equivalent to regression calibration for that special case. Basically, one replaces error-prone covariates by the conditional expectations of true covariates

given error-prone measurements and other covariates measured without error and then proceeds with standard estimation techniques.

How regression calibration is implemented in the matched study setting is not entirely obvious. The natural covariate vectors in this setting are the control minus case covariate difference vectors. In our first attempt to use regression calibration for this problem, we applied regression calibration to these difference vectors. We will refer to this version as regression calibration method 1. Also, we considered a second version, which proceeds exactly as in the unmatched setting, ignoring correlation among members of a matched pair. We shall refer to this as regression calibration method 2.

For regression calibration method 1, we derive an estimate of $E[d_{kx} | d_{kz}, d_{kw}]$. Conditional on the error-free covariates $\{z_{kj}^{(i)}\}$, we model d_{kx} and d_{kw} as $d'_{kx} = \Lambda_0 + d'_{kz} \Lambda_1 + d'_{k\epsilon}$ and $d'_{kw} = d'_{kx} + d'_{ku} = \Lambda_0 + d'_{kz} \Lambda_1 + d'_{k\epsilon} + d'_{ku}$, $k = 1, 2, \dots, K$, where Λ_0 and Λ_1 are unknown coefficient matrices and $d_{k\epsilon}$ and d_{ku} are zero-mean multivariate error variables. Define $d_{k\eta} = d_{k\epsilon} + d_{ku}$, $\Sigma_{d_\eta, d_\eta} = \text{cov}(d_{k\eta}, d_{k\eta})$, and $\Sigma_{d_\epsilon, d_\epsilon} = \text{cov}(d_{k\epsilon}, d_{k\epsilon})$. Assume a multivariate regression model (Johnson and Wichern, 1988) $D_w = D_z \Lambda + D_\eta$, where D_w is a $K \times p_1 M$ matrix with k th row equal to d'_{kw} , D_η is a $K \times p_1 M$ matrix with k th row equal to $d'_{k\eta}$, Λ is a $(p_2 M + 1) \times p_1 M$ matrix with the first row equal to Λ_0 and the remaining portion equal to Λ_1 , and D_z is a $K \times (p_2 M + 1)$ matrix consisting of a $K \times p_2 M$ matrix with the k th row equal to d'_{kz} , augmented by a leading column of ones. This yields ordinary least squares estimates $\hat{\Lambda} = (D_z' D_z)^{-1} D_z' D_w$, $\hat{D}_\eta = D_w - D_z \hat{\Lambda}$, and $\hat{\Sigma}_{d_\eta, d_\eta} = \hat{D}_\eta' \hat{D}_\eta / [K - (p_2 M + 1)]$.

A consistent estimator of $\Sigma_{d_\epsilon, d_\epsilon}$ is $\hat{\Sigma}_{d_\epsilon, d_\epsilon} = \hat{\Sigma}_{d_\eta, d_\eta} - \hat{\Sigma}_{d_u, d_u}$, where $\hat{\Sigma}_{d_u, d_u}$ is a consistent estimate of Σ_{d_u, d_u} . The best linear approximation (exact under multivariate normality) to $E[d_{kx} | d_{kz}, d_{kw}]$ is

$$\begin{aligned} & E[d_{kx} | d_{kz}] + \text{cov}(d_{kx}, d_{kw} | d_{kz}) [\text{cov}(d_{kw}, d_{kw} | d_{kz})]^{-1} \\ & \times [d_{kw} - E[d_{kw} | d_{kz}]] \\ & = (I - \Sigma_{d_\epsilon, d_\epsilon} \Sigma_{d_\eta, d_\eta}^{-1}) (\Lambda_0 + d'_{kz} \Lambda_1) \\ & + \Sigma_{d_\epsilon, d_\epsilon} \Sigma_{d_\eta, d_\eta}^{-1} d_{kw}. \end{aligned}$$

Therefore, we estimate $E[d_{kx} | d_{kz}, d_{kw}]$ by

$$(I - \hat{\Sigma}_{d_\epsilon, d_\epsilon} \hat{\Sigma}_{d_\eta, d_\eta}^{-1}) (\hat{\Lambda}_0 + d'_{kz} \hat{\Lambda}_1) + \hat{\Sigma}_{d_\epsilon, d_\epsilon} \hat{\Sigma}_{d_\eta, d_\eta}^{-1} d_{kw},$$

$$k = 1, 2, \dots, K,$$

and substitute into the likelihood (3) to perform the usual analysis to estimate β .

Regression calibration method 2 is implemented as in the unmatched setting. The w_{kj} , z_{kj} , and $\Sigma_{u,u}$ replace d_{kw} , d_{kz} , and Σ_{d_u, d_u} in the above formulas; i.e., $\{(w_{kj}, z_{kj})\}$ are treated as $K(M+1)$ independent p -dimensional covariate vectors, ignoring all matching. The matrix $\Sigma_{u,u}$ is $p_1 \times p_1$ with i th diagonal element equal to $\sigma_O(i, i) + \sigma_B^2(i) + \sigma_E^2(i)$ and (i, i') element equal to $\sigma_O(i, i')$.

6. Example

Several recent studies (Barrett-Connor et al., 1990; Hsing and Comstock, 1993; Gann et al., 1996; Nomura et al., 1996) have examined the association of hormones with prostate cancer

risk. Findings differed, perhaps partly due to noise inherent in hormone measures, including laboratory variability, variations in specimen collection procedures, and random biological fluctuations over time (Hsing, 1996).

As part of the Alpha-Tocopherol Beta-Carotene (ATBC) Lung Cancer Prevention Study (ATBC Cancer Prevention Study Group, 1994), a serum repository was created, and several prospective matched case-control studies have been conducted. Our example examines the relationship between serum sex hormones and prostate cancer. As the 29,133 male smokers were accrued, serum samples were collected and frozen at -70°C and sociodemographic and anthropometric variables were recorded. After 5–8 years of follow-up, 246 men had developed prostate cancer and 116 were randomly selected for inclusion in the hormone study. For each case, we identified two controls who were free of prostate cancer at the time of the case's diagnosis and matched on the basis of clinic, treatment group, age at time of case's diagnosis (± 1 year, relaxed to ± 2 years in a few cases), and date of blood draw (± 28 days, relaxed to ± 45 days for a few cases). The 111 case-control sets with no missing data on the variables of interest were included in this analysis. Serum samples from those 333 men were thawed and assayed for a battery of nine hormones and binding proteins.

We present results from fitting a model that includes testosterone and dihydrotestosterone (DHT, a metabolite of testosterone) and the variables educational status (categorized as common school only, some high school, or high school graduate) and height (in centimeters) as examples of sociodemographic and anthropometric variables. For the k th matched set, the assumed risk model is

$$\begin{aligned} & \text{logit}\{\text{Pr}(Y = 1 | k, T, D, H, Z_1, Z_2)\} \\ & = a_k + \beta_T T + \beta_D D + \beta_H H + \beta_1 Z_1 + \beta_2 Z_2, \end{aligned}$$

where Y = binary indicator of prostate cancer, a_k = contribution of matching variables in k th stratum, T = $\log(\text{testosterone})$, D = $\log(\text{dihydrotestosterone})$, H = height in centimeters, and Z_1 and Z_2 = binary indicators of educational levels 2 and 3, respectively.

A separate study was conducted to assess the magnitude of laboratory and occasion-within-person variability in serum testosterone and DHT measurements. Men participating in this study were in the same age range (50–69 years), but not all were smokers as in the ATBC cohort. Twenty-three men completed at least four of the six scheduled blood draws, two vials per draw. The resulting 262 vials were randomly distributed among 12 assay batches for DHT and 15 batches for testosterone. Measurement error variance estimates (SE) were $\text{var}(\epsilon_{kj}^{(1)}) = .0038$ (.00049), $\text{var}(B_k^{(1)}) = .0039$ (.0016), $\text{var}(O_{kj}^{(1)}) = .026$ (.0046) for testosterone, $\text{var}(\epsilon_{kj}^{(2)}) = .0077$ (.00098), $\text{var}(B_k^{(2)}) = .0037$ (.0021), $\text{var}(O_{kj}^{(2)}) = .032$ (.0061) for DHT, and $\text{cov}(O_{kj}^{(1)}, O_{kj}^{(2)}) = .024$ (.0047).

Table 1 presents the results of fitting the logistic risk model using no correction for measurement error (naive analysis), using a conditional scores adjustment, and using two versions of regression calibration. The adjustments (assuming known measurement error) resulted in substantial corrections to coefficients of the error-prone covariates but little for the error-free covariates. All results in Table 1 were calculated using

Table 1

Uncorrected and measurement error corrected analyses for hormones and prostate cancer example under the risk model $\text{logit}[\text{Pr}(Y = 1 | A, T, D, H, Z_1, Z_2)] = A + \beta_T T + \beta_D D + \beta_H H + \beta_1 Z_1 + \beta_2 Z_2$

Variable	Estimated coefficient	Standard error ^a	Approximate z-score ^b	p-value
a. Uncorrected (Naive) Analysis				
log testosterone (<i>T</i>)	.4399	.5940	.74	.46
log DHT (<i>D</i>)	-.6019	.5591	-1.08	.28
Height in cm (<i>H</i>)	-.0245	.0182	-1.35	.18
Educational status				
<i>Z</i> ₁	.1465	.3235	.45	.65
<i>Z</i> ₂	.4904	.4165	1.18	.24
b. Conditional Scores Corrected Analysis Assuming Measurement Error Variances Are Known				
log testosterone (<i>T</i>)	.8836	1.1986	.74	.46
log DHT (<i>D</i>)	-1.1404	1.2045	-.95	.34
Height in cm (<i>H</i>)	-.0250	.0173	-1.45	.15
Educational status				
<i>Z</i> ₁	.1358	.3606	.38	.71
<i>Z</i> ₂	.5046	.4422	1.14	.25
c. Regression Calibration Method 1 Assuming Measurement Error Variances Are Known				
log testosterone (<i>T</i>)	1.4085	1.8012	.78	.43
log DHT (<i>D</i>)	-1.7792	1.8850	-.94	.35
Height in cm (<i>H</i>)	-.0253	.0176	-1.44	.15
Educational status				
<i>Z</i> ₁	.1141	.3658	.31	.76
<i>Z</i> ₂	.5124	.4545	1.13	.26
d. Regression Calibration Method 2 Assuming Measurement Error Variances Are Known				
log testosterone (<i>T</i>)	.7294	.9905	.74	.46
log DHT (<i>D</i>)	-.9601	.9588	-1.00	.32
Height in cm (<i>H</i>)	-.0243	.0170	-1.42	.15
Educational status				
<i>Z</i> ₁	.1384	.3603	.38	.70
<i>Z</i> ₂	.4987	.4374	1.14	.25

^a Standard errors computed for the uncorrected analysis (part a) are the maximum likelihood-based estimates. For the analyses in parts b-d, standard errors are jackknife estimates based on (6) setting measurement error variances equal to estimated values.

^b $z = \text{estimate}/(\text{SE})$.

a Fortran program we developed and took 5 seconds on a SGI Power Challenge supercomputer. Also available from the authors is an SAS macro that calls PROC PHREG (SAS Institute, 1996) iteratively to estimate the conditional scores parameter estimates and their standard errors when error variances are known. Using the SAS macro, calculations in Table 1, part b, took 23 minutes on a Sun Workstation.

To account for the variability in the measurement error variance component estimates, we also performed a parametric bootstrap procedure in which we simulated 500 sets of measurement error variance estimates using the estimated asymptotic multivariate Gaussian sampling distribution of the estimates obtained from our variability study. For each of N simulated sets of measurement error variance estimates, we apply the correction procedures and compute the jackknife variance estimate. Then the final standard error estimate of

the j th element of the beta vector is the square root of

$$\widehat{\text{var}}_{\text{bootjack}}(\hat{\beta}_j) = \frac{\sum_{i=1}^N (\hat{\beta}_j^{(i)} - \hat{\beta}_j^{(\cdot)})^2}{(N-1)} + \frac{\sum_{i=1}^N \widehat{\text{var}}_{\text{jack}}(\hat{\beta}_j^{(i)})}{N}, \quad (7)$$

where $\hat{\beta}_j^{(i)}$ is the corrected estimate of the j th element of β computed using the full collection of matched sets and assuming measurement error variance components equal to the i th simulated set, $\hat{\beta}_j^{(\cdot)}$ is the mean of those N values, and $\widehat{\text{var}}_{\text{jack}}(\hat{\beta}_j^{(i)})$ is the j th diagonal element of (6) computed assuming measurement error variances equal to the i th simulated set. This follows from $\text{var}(\hat{\beta}) = \text{var}[E(\hat{\beta} | \hat{\Sigma}_{d_u, d_u})] + E[\text{var}(\hat{\beta} | \hat{\Sigma}_{d_u, d_u})]$. Due to the very precise estimation of our measurement error variances, adding the parametric boot-

strap resulted in minimal changes in standard errors and test statistics (not presented).

7. Simulation Studies

7.1 Simulation Design

Case-control data sets of $K = 111$ and 300 matched sets were simulated under a risk model estimated from the prostate cancer example and under some variations, with 1000 repetitions per study. Measurement error variances were assumed known.

Covariate vectors (a_k, T, D, H, Z_1, Z_2) and responses, Y , were generated for a large cohort of subjects by randomly generating multivariate normal vectors and applying appropriate transformations or cutpoints. (Further details available from the authors.) The binary outcome variable Y was generated under the model

$$\begin{aligned} \text{logit}[\text{Pr}(Y = 1 \mid A, T, D, H, Z_1, Z_2)] \\ = A + \beta_T T + \beta_D D + \beta_H H + \beta_1 Z_1 + \beta_2 Z_2. \end{aligned}$$

Tables 2 and 3 simulations use parameters estimated in Table 1, part b. Those values for β_T , β_D , and β_H translate to relative risks of 1.4, 0.61, and 0.80, respectively, comparing the third quartile of the covariate distribution with the first. The β_1 and β_2 translate to relative risks of 1.2 and 1.6, respectively, comparing with lowest educational level. Using A , controls were matched 2:1 to cases to form K sets. Error-prone covariates T_e and D_e were generated using measurement error model (2) with estimates from our variability study.

Naive estimates were obtained by conditional logistic regression analysis using T_e and D_e . Conditional scores estimates were obtained as described in Section 4. Regression calibration method 1 and 2 estimates were obtained as described in Section 5. Naive standard error estimates were the usual ones based on the information matrix supplied by conditional logistic regression routines (e.g., SAS PROC PHREG; SAS Institute, 1996). Jackknife standard error estimates were used in the other cases.

7.2 Simulation Results Under the Risk Model Estimated from the Prostate Cancer Example

All methods converged on all repetitions of the simulations in Table 2 ($K = 111$) and Table 3 ($K = 300$). As expected, the naive analysis produced severely attenuated estimates for β_T and β_D . Regression calibration method 1 also produced severely biased estimates; hence, we consider it no further. Conditional scores showed a little more small-sample bias than regression calibration method 2, but bias in both cases was fairly small for 300 or more matched sets. Compared with no correction, both correction methods produced estimates with larger root mean-squared errors and median absolute errors in small samples; but when the number of matched sets increased to 300 or 500 (results not shown), the bias dominated, and this reversed. Using conditional scores or regression calibration method 2, interval coverages were close to nominal levels. Naive estimation resulted in severe undercoverage. Standard error estimates were essentially unbiased.

A simulation for a null ($\beta_T = 0$) case (results not shown) demonstrated that all methods have an approximately correct level, as predicted by results of Tosteson and Tsiatis (1988) and Carroll et al. (1995, Section 11.4).

7.3 Simulation Results Under Varied Risk and Measurement Error Structures

Bias and convergence problems were noted in exploratory studies examining the effects of factors such as degree of correlation among covariates, departures from Gaussian errors, large relative risks, large measurement error, and skewness of covariate distributions. In each situation, we simulated 1000 data sets of 111 matched sets.

Increasing the correlation between H and T and between H and D from -0.1 to $+0.7$ increased the degree of the measurement error correction on β_H to about 25%. The bias increased only slightly, and the standard deviation nearly doubled. Thus, the minimal adjustment on the error free covariate coefficients in our original example was likely due to the small correlations between variables measured with and without error.

To examine the effects of non-Gaussian measurement error, we generated errors as Gaussian with point masses at ± 3 standard deviations, occurring with probability .05 at each tail. Bias was almost 30% for β_T and a little more than 20% for β_D for both conditional scores and regression calibration method 2.

To simulate large relative risks, we multiplied both β_T and β_D by three. Conditional scores estimates were only slightly more biased (compare to Table 2, part b), but convergence failed in 7 of 1000 simulation repetitions.

To examine the effect of large measurement error, measurement error was increased threefold. The conditional scores procedure converged in only 784 of 1000 repetitions, whereas regression calibration always converged. Bias was small in both cases.

A key comparison between conditional scores and regression calibration involves covariates from non-Gaussian distributions. Calibration function linearity is satisfied for multivariate Gaussian covariates, so one might expect regression calibration to perform best in this case and worse under departures such as highly skewed distributions. In contrast, conditional scores are independent of the true covariate distributions. We simulated covariates T and D as log normal rather than normal, with means as before but with variances now set equal to means. Measurement error variances were set to 40% of the true covariate variances to maintain the original relative proportion of measurement error.

Results for the highly skewed covariate situation are presented in Table 4. Conditional scores estimates were less biased than the regression calibration estimates, although the former were more variable and their root mean-squared errors and median absolute errors were larger. But the conditional scores procedure converged on only 781/1000 data sets. Convergence was more problematic (475/1000) for calculation of the jackknife standard error, as it required convergence on all jackknife samples.

Simulating data sets of 300 matched sets (results not shown), the conditional scores procedure converged in 954/1000 repetitions and the jackknife standard error could be computed in 888/1000 repetitions. The biases of the conditional scores estimates of β_T and β_D were reduced to less than 6%. Biases in the regression calibration method 2 estimates remained high at 15% and 20%, respectively, and resulted in gross undercoverage of confidence intervals.

Table 2

Simulation results based on 1000 repetitions when simulating and fitting the risk model $\text{logit}[\Pr(Y = 1 | A, T, D, H, Z_1, Z_2)] = A + \beta_T T + \beta_D D + \beta_H H + \beta_1 Z_1 + \beta_2 Z_2$ with $K = 111$ matched sets

Coefficient:	β_T	β_D	β_H	β_1	β_2
True value:	.88	-1.14	-.025	.14	.50
a. Naive Analysis					
Average estimate	.53	-.73	-.026	.13	.51
Percent bias	-40.0	36.0	-2.9	-4.2	2.9
Average SE estimate ^a	.61	.53	.018	.33	.40
Monte Carlo SE	.61	.53	.018	.33	.41
(Mean-squared error) ^{1/2}	.71	.67	.018	.33	.41
Median absolute error	.49	.49	.012	.22	.26
Coverage probability (in %) of nominal ^b					
90% interval	82.7	80.0	90.5	88.8	89.6
95% interval	91.1	87.4	95.7	93.9	94.7
b. Conditional Scores Analysis					
Average estimate	.96	-1.24	-.027	.13	.51
Percent bias	8.9	-8.4	-7.1	-8.4	1.7
Average SE estimate ^c	1.14	1.00	.020	.35	.44
Monte Carlo SE	1.09	.95	.019	.34	.42
(Mean-squared error) ^{1/2}	1.09	.95	.019	.34	.42
Median absolute error	.70	.59	.013	.22	.28
Coverage probability (in %) of nominal ^b					
90% interval	92.9	93.1	92.1	90.8	91.9
95% interval	97.3	96.8	96.3	95.2	96.1
c. Regression Calibration Method 1					
Average estimate	1.49	-1.85	-.028	.12	.50
Percent bias	69.8	-62.7	-11.6	-13.1	.4
Average SE estimate ^c	1.78	1.55	.021	.37	.47
Monte Carlo SE	1.67	1.46	.020	.35	.44
(Mean-squared error) ^{1/2}	1.78	1.63	.020	.35	.44
Median absolute error	1.11	1.00	.013	.22	.29
Coverage probability (in %) of nominal ^b					
90% interval	92.2	92.1	92.5	91.4	92.2
95% interval	96.9	96.9	96.5	95.5	96.2
d. Regression Calibration Method 2					
Average estimate	.93	-1.19	-.026	.13	.51
Percent bias	5.2	-4.2	-4.5	-6.5	1.5
Average SE estimate ^c	1.06	.92	.019	.34	.43
Monte Carlo SE	1.02	.89	.018	.33	.42
(Mean-squared error) ^{1/2}	1.02	.89	.018	.33	.42
Median absolute error	.67	.58	.012	.22	.27
Coverage probability (in %) of nominal ^b					
90% interval	91.9	91.2	91.7	90.2	91.8
95% interval	96.4	96.0	96.1	95.3	95.8

^a Square-root of mean maximum likelihood-based variance estimate.

^b Interval computed as estimate $\pm z_{\alpha/2}$ SE, where $z_{\alpha/2}$ is the appropriate standard normal percentage point.

^c Square-root of mean jackknife variance estimate computed using (6).

8. Discussion

Both the conditional scores and regression calibration method 2 measurement error adjustments performed well for bias correction in our simulated examples involving true covariates generated from Gaussian distributions, moderate relative risks, and moderate Gaussian measurement error. Only slightly more bias and variability were observed in the conditional

scores estimates compared with the regression calibration method 2 estimates in these well-behaved settings. One would expect regression calibration to have some advantage when true covariates are Gaussian because the form of the assumed calibration function is exactly correct when all true covariates are Gaussian. Indeed, the conditional scores procedure produced much less biased estimates when true covariate distributions

Table 3

Simulation results based on 1000 repetitions when simulating and fitting the risk model $\text{logit}[\text{Pr}(Y = 1 | A, T, D, H, Z_1, Z_2)] = A + \beta_T T + \beta_D D + \beta_H H + \beta_1 Z_1 + \beta_2 Z_2$ with $K = 300$ matched sets

Coefficient:	β_T	β_D	β_H	β_1	β_2
True value:	.88	-1.14	-.025	.14	.50
a. Naive Analysis					
Average estimate	.52	-.70	-.025	.14	.51
Percent bias	-40.8	38.6	1.4	1.8	1.0
Average SE estimate ^a	.36	.32	.011	.19	.24
Monte Carlo SE	.37	.31	.011	.20	.23
(Mean-squared error) ^{1/2}	.51	.54	.011	.20	.23
Median absolute error	.38	.44	.0077	.13	.15
Coverage probability (in %) of nominal ^b					
90% interval	74.0	59.1	89.9	88.6	91.6
95% interval	83.0	71.4	94.3	93.9	95.6
b. Conditional Scores Analysis					
Average estimate	.92	-1.16	-.025	.14	.50
Percent bias	4.0	-1.5	-1.5	-1.3	-.5
Average SE estimate ^c	.63	.55	.011	.20	.25
Monte Carlo SE	.62	.53	.011	.20	.24
(Mean-squared error) ^{1/2}	.62	.53	.011	.20	.24
Median absolute error	.41	.35	.0079	.13	.15
Coverage probability (in %) of nominal ^b					
90% interval	89.9	92.2	90.0	89.5	91.7
95% interval	95.0	96.0	95.0	94.2	96.1
c. Regression Calibration Method 1					
Average estimate	1.52	-1.84	-.026	.13	.49
Percent bias	73.0	-61.3	-5.4	-5.2	-2.4
Average SE estimate ^c	1.03	.90	.012	.21	.26
Monte Carlo SE	1.00	.86	.012	.21	.24
(Mean-squared error) ^{1/2}	1.19	1.11	.012	.21	.24
Median absolute error	.78	.78	.0082	.13	.15
Coverage probability (in %) of nominal ^b					
90% interval	85.5	83.2	90.1	89.8	91.2
95% interval	92.5	91.4	94.9	94.4	96.3
d. Regression Calibration Method 2					
Average estimate	.91	-1.14	-.025	.14	.50
Percent bias	2.9	.2	.1	-.5	-.4
Average SE estimate ^c	.61	.53	.011	.20	.24
Monte Carlo SE	.60	.51	.011	.20	.23
(Mean-squared error) ^{1/2}	.60	.51	.011	.20	.23
Median absolute error	.40	.34	.0078	.13	.15
Coverage probability (in %) of nominal ^b					
90% interval	90.0	91.5	90.0	89.2	91.8
95% interval	94.6	95.8	94.9	94.1	96.1

^a Square-root of mean maximum likelihood-based variance estimate.

^b Interval computed as estimate $\pm z_{\alpha/2}$ SE, where $z_{\alpha/2}$ is the appropriate standard normal percentage point.

^c Square-root of mean jackknife variance estimate computed using (6).

were non-Gaussian and highly skewed. Also, the greater variability in the conditional scores estimates likely reflects some loss of efficiency resulting from requiring no assumptions about the true covariate distributions.

Regression calibration method 1, based on case-control differences, performed very poorly even in the Gaussian covariate setting and should not be considered a viable option.

Based on the dependence of the likelihood on control-case covariate differences, this at first seemed the most natural way to apply regression calibration in this matched setting. Perhaps its failure was due to its implicitly requiring information about case/control status in order to form the case-control differences of the covariates. The calibration step of regression calibration should not require this knowledge of outcome.

Table 4

Simulation results based on 1000 repetitions when simulating and fitting the risk model
 $\text{logit}[\Pr(Y = 1 | A, T, D, H, Z_1, Z_2)] = A + \beta_T T + \beta_D D + \beta_H H + \beta_1 Z_1 + \beta_2 Z_2$
 with $K = 111$ matched sets and highly skewed covariate distributions^a

Coefficient:	β_T	β_D	β_H	β_1	β_2
True value:	.88	-1.14	-.025	.14	.50
a. Conditional Scores Analysis (Converged in 781/1000 Repetitions)					
Average estimate	.94	-1.22	-.027	.10	.49
Percent bias	6.8	-6.9	-7.1	-27.0	-2.7
Average SE estimate ^b	.31	.47	.034	.62	.81
Monte Carlo SE	.25	.38	.032	.59	.78
(Mean-squared error) ^{1/2}	.26	.39	.032	.59	.78
Median absolute error	.14	.21	.020	.37	.49
Coverage probability (in %) of nominal ^c					
90% interval	95.1	94.9	98.1	98.5	96.9
95% interval	97.4	96.4	99.6	99.4	98.6
b. Regression Calibration Method 2 (Converged in 1000/1000 Repetitions)					
Average estimate	.77	-.94	-.022	.11	.44
Percent bias	-12.3	17.5	11.0	-22.6	-12.8
Average SE estimate ^b	.14	.19	.024	.42	.55
Monte Carlo SE	.12	.18	.022	.41	.53
(Mean-squared error) ^{1/2}	.16	.27	.022	.41	.54
Median absolute error	.13	.22	.015	.27	.35
Coverage probability (in %) of nominal ^c					
90% interval	73.7	66.7	92.4	91.3	91.3
95% interval	82.7	77.3	97.2	96.4	95.7

^a T and D were simulated as log normal with $E(T) = 6.37$, $SD(T) = 2.5$, $E(D) = 3.97$, $SD(D) = 2.0$. Measurement error variances were 40% of the true covariate variances.

^b Square-root of mean jackknife variance estimate computed using (6).

^c Interval computed as estimate $\pm z_{\alpha/2}SE$, where $z_{\alpha/2}$ is the appropriate standard normal percentage point.

The approach of Armstrong et al. (1989) explicitly uses the case/control information as part of a discriminant analysis model and therefore may be a good choice if one is willing to make the necessary multivariate normal discriminant model assumptions.

If one uses root mean-squared error as a summary combined measure of bias and variability, then neither the conditional scores nor regression calibration method 2 measurement error correction confers an advantage over no correction for small numbers of matched sets. But as the number of matched sets increases, both correction procedures produce mean-squared errors that are the same or better (smaller) than those obtained using no correction. Furthermore, if we use confidence interval coverage as our criterion, then the correction methods produce uniformly better results.

When relative risks or measurement errors were large but covariate distributions were still Gaussian, the conditional scores procedure was prone to convergence problems and resulted in slightly more biased estimates in small samples. Non-Gaussian measurement error caused both correction methods to produce biased estimates. When true covariates were generated from non-Gaussian highly skewed distributions, regression calibration method 2 produced substantially biased estimates. Conditional scores experienced some convergence problems in these settings, but when it did converge, it produced estimates with smaller bias and much better confidence

interval coverage. Moreover, the convergence problems could be greatly reduced by increasing the number of matched sets. Thus, the settings in which the conditional scores method would be most useful and preferred over the regression calibration method include those in which there is at least a moderately large number of matched sets and it is suspected that true covariate distributions may be highly skewed. Although sometimes highly skewed covariates can be transformed to Gaussian, this is often not possible, and transformations also can make model interpretation more difficult.

Our simulations assumed that the true measurement error variances were known. More likely, only consistent estimates of the measurement error variance components would be available, and the variability in them would introduce additional variability into the corrected parameter estimates. One would at least want to perform a sensitivity analysis by varying the assumed degree of measurement error. As a better alternative, we described a parametric bootstrapping procedure using the estimated asymptotic distribution of the variance estimates to adjust for variability in measurement error variance estimates. For small variability studies, the appropriateness of the asymptotic distribution might be questionable. One could bootstrap the entire variability study data set and reestimate the measurement error variances for each data set, but this could be computationally prohibitive. If one estimates measurement error variances from an external variability study,

great care also must be taken so that the variability characteristics of the external study are representative of the variability in the main study. The dramatic effect demonstrated in this study of the measurement error on bias of the parameter estimates should serve as motivation for researchers to collect sufficient samples to properly estimate measurement error as part of their main study. In many cases, when biorepositories are initiated, it would be feasible to plan to collect multiple specimens over time from a subset of study participants in order to obtain measurement error variance estimates from an internal variability study.

Several generalizations of the conditional scores method are possible. Our derivation assumed Gaussian nondifferential measurement error with constant variance. Problems of heteroscedasticity may potentially be handled by appropriate transformations of the error-prone covariates. Non-Gaussian measurement error would result in a different form for the sufficient statistics for the $d_{k,x}$'s, but the sufficient statistics may be difficult to derive. The method can be easily generalized to unequal matching. Similarly, our methods could be adapted to handle replicated w 's on some subjects, such as from an internal variability study. If samples within a matched set are not matched on batch, we discussed how the derivation of the sufficient statistics could become difficult. However, if the batch-to-batch variance is very small relative to the other measurement error variance components, it may be possible to ignore the batch effects and the correlations they induce and still obtain reasonable corrected estimates.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the Alpha-Tocopherol Beta-Carotene Cancer Prevention Study (supported by NCI Public Health Service contract N01CN45165 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services) for use of the data. R. Carroll's research was supported by a grant from the National Cancer Institute (CA-57030) and was partially completed during a visit to the Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute. This work was initiated while L. McShane was a member of the Biometry Branch, Division of Cancer Prevention and Control, National Cancer Institute.

RÉSUMÉ

Nous proposons une méthode basée sur des scores conditionnels pour obtenir des estimateurs, corrigés pour le biais, des log odds ratios dans les études cas-contrôles appariées, où une ou plusieurs covariables sont sujettes à erreurs de mesure. L'approche suppose que l'on conditionne par rapport à des statistiques exhaustives pour les valeurs exactes non observables des covariables, celles-ci étant traitées comme des paramètres fixes inconnus. Dans le cas d'erreurs de mesure gaussiennes non-différentiables, nous obtenons un ensemble d'équations de scores fidèles (non biaisés) permettant d'estimer le log OR des paramètres étudiés. La procédure permet d'éliminer avec succès le biais des estimations naïves, et les erreurs types des estimations sont obtenues par des méthodes de rééchantillonnage. Nous présentons un exemple appliqué à des données d'une étude cas-contrôle appariée du cancer de la prostate et des taux d'hormone sériques circulants, et nous comparons la performance de notre méthode avec celle des procédures de calibration par régression.

REFERENCES

- Armstrong, B. G., Whittemore, A. S., and Howe, G. R. (1989). Analysis of case-control data with covariate measurement error: Application to diet and colon cancer. *Statistics in Medicine* 8, 1151-1163.
- ATBC Cancer Prevention Study Group. (1994). The Alpha-Tocopherol, Beta-Carotene Lung Cancer Prevention Study: Design, methods, participant characteristics, and compliance. *Annals of Epidemiology* 4, 1-10.
- Barrett-Connor, E., Garland, C., McPhillips, J. B., Khaw, K. T., and Wingard, D. L. (1990). A prospective population-based study of androstenedione, estrogens, and prostatic cancer. *Cancer Research* 50, 169-173.
- Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*. London: Chapman and Hall.
- Forbes, A. B. and Santner, T. J. (1995). Estimators of the odds ratio regression parameters in matched case-control studies with covariate measurement error. *Journal of the American Statistical Association* 90, 1075-1084.
- Gann, P. H., Hennekens, C. H., Ma, J., Longcope, C., and Stampfer, M. J. (1996). Prospective study of sex hormone levels and risk of prostate cancer. *Journal of the National Cancer Institute* 88, 1118-1126.
- Hosmer, D. W., Jr., and Lemeshow, S. (1989). *Applied Logistic Regression*. New York: Wiley.
- Hsing, A. W. (1996). Hormones and prostate cancer: Where do we go from here? (editorial). *Journal of the National Cancer Institute* 88, 1093-1095.
- Hsing, A. W. and Comstock, G. W. (1993). Serological precursors of cancer: Serum hormones and the risk of subsequent prostate cancer. *Cancer Epidemiology, Biomarkers, and Prevention* 2, 27-32.
- Johnson, R. A. and Wichern, D. W. (1988). *Applied Multivariate Statistical Analysis*, 2nd edition. Englewood Cliffs, New Jersey: Prentice-Hall.
- Nomura, A. M., Stemmermann, G. N., Chyou, P. H., and Stanczyk, F. Z. (1996). Serum androgens and prostate cancer. *Cancer Epidemiology, Biomarkers, and Prevention* 5, 621-625.
- Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* 69, 331-342.
- Rosner, B., Willett, W. C., and Spiegelman, D. (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine* 8, 1051-1069.
- Rosner, B., Spiegelman, D., and Willett, W. C. (1990). Correction of logistic regression relative risk estimates and confidence intervals for measurement error: The case of multiple covariates measured with error. *American Journal of Epidemiology* 132, 734-745.
- Rosner, B., Spiegelman, D., and Willett, W. C. (1992). Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. *American Journal of Epidemiology* 136, 1400-1413.

SAS Institute. (1996). The PHREG procedure. In *SAS/STAT Software: Changes and Enhancements through Release 6.11*, 809-884. Cary, North Carolina: SAS Institute.

Stefanski, L. A. and Carroll, R. J. (1987). Conditional scores and optimal scores for generalized measurement-error models. *Biometrika* 74, 703-716.

Tosteson, T. and Tsiatis, A. (1988). The asymptotic relative efficiency of score tests in a generalized linear model with surrogate covariates. *Biometrika* 75, 507-514.

Received June 1998. Revised March and May 2000.

Accepted June 2000.

APPENDIX A

Derivation of Conditional Likelihood

Observe that the likelihood (4) equals $\prod_{k=1}^K l_k(\beta)$, where $l_k(\beta) = \Pr\{Y_k | x_k, z_k, (T_k = 1)\}$, allowing us to perform some algebraic manipulations on $l_k(\beta)$ to put it in a form more amenable to deriving a sufficient statistic. Noting that

$$-(x'_{k1}\beta_x + z'_{k1}\beta_z) = -\sum_{j=1}^{M+1} Y_{kj} (x'_{k1}\beta_x + z'_{k1}\beta_z)$$

when $T_k = \sum_{j=1}^{M+1} Y_{kj} = 1$ and multiplying numerator and denominator of $l_k(\beta)$ by $\exp[-(x'_{k1}\beta_x + z'_{k1}\beta_z)]$, we obtain

$$l_k(\beta) = \frac{\exp\left[\sum_{j=2}^{M+1} Y_{kj} \{(x_{kj} - x_{k1})'\beta_x + (z_{kj} - z_{k1})'\beta_z\}\right]}{1 + \sum_{j=2}^{M+1} \exp\{(x_{kj} - x_{k1})'\beta_x + (z_{kj} - z_{k1})'\beta_z\}}$$

Write the denominator as $\{S_1(d_{kx}, d_{kz}, \beta)\}^{-1}$. Define $B_{kx} = (Y_{k2}\beta'_x, Y_{k3}\beta'_x, \dots, Y_{k,M+1}\beta'_x)'$ and $B_{kz} = (Y_{k2}\beta'_z, Y_{k3}\beta'_z, \dots, Y_{k,M+1}\beta'_z)'$ so that

$$l_k(\beta) = S_1(d_{kx}, d_{kz}, \beta) \exp(B'_{kx}d_{kx} + B'_{kz}d_{kz}) \\ = \Pr\{Y_k | x_k, z_k, (T_k = 1)\}.$$

Now we show that $\Delta_k = d_{kw} + \Sigma_{d_u, d_u} B_{kx}$ is sufficient for d_{kx} , treating β_x as known, when the $\{d_{ku}\}$ are independent across matched sets. Sufficiency is demonstrated by showing that $\Pr\{Y_k | \Delta_k, d_{kx}, d_{kz}, (T_k = 1)\}$ does not depend on d_{kx} . The independence of the $\{d_{ku}\}$ allows one to derive the sufficient statistics separately on each $l_k(\beta)$. This independence is achieved under our measurement error model when all samples from the same matched set are assayed in the same lab batch. If the $\{d_{ku}\}$ (and hence the $\{d_{kw}\}$) were dependent, then sufficient statistics would have to be derived by starting with the full likelihood $\prod_{k=1}^K l_k(\beta)$ and conditioning on a joint distribution of sufficient statistics. This makes the derivation considerably more difficult, sometimes intractable. Under Gaussian nondifferential measurement error and independence of the $\{d_{ku}\}$ across strata, $\Pr\{Y_k | d_{kw}, d_{kx}, d_{kz}, (T_k = 1)\} = \Pr\{Y_k | d_{kx}, d_{kz}, (T_k = 1)\}$, so

$$\Pr\{Y_k, d_{kw} | d_{kx}, d_{kz}, (T_k = 1)\} \\ = \Pr\{Y_k | d_{kx}, d_{kz}, (T_k = 1)\}$$

$$\times \Pr\{d_{kw} | d_{kx}, d_{kz}, (T_k = 1)\} \\ = \text{constant} \times l_k(\beta) \\ \times \exp\left\{-\frac{1}{2}(d_{kw} - d_{kx})'\Sigma_{d_u, d_u}^{-1}(d_{kw} - d_{kx})\right\} \\ = S_2(d_{kx}, d_{kz}, \beta) \\ \times \exp\left\{(d_{kw} + \Sigma_{d_u, d_u} B_{kx})'\Sigma_{d_u, d_u}^{-1}d_{kx} + B'_{kz}d_{kz} \right. \\ \left. - \frac{1}{2}d'_{kw}\Sigma_{d_u, d_u}^{-1}d_{kw}\right\},$$

where

$$S_2(d_{kx}, d_{kz}, \beta) = \text{constant} \times S_1(d_{kx}, d_{kz}, \beta) \\ \times \exp\left(-\frac{1}{2}d'_{kx}\Sigma_{d_u, d_u}^{-1}d_{kx}\right).$$

Transforming from (Y_k, d_{kw}) to (Y_k, Δ_k) gives

$$\Pr\{Y_k, \Delta_k | d_{kx}, d_{kz}, (T_k = 1)\} \\ = S_2(d_{kx}, d_{kz}, \beta) \\ \times \exp\left(\Delta'_k \Sigma_{d_u, d_u}^{-1}d_{kx} - \frac{1}{2}\Delta'_k \Sigma_{d_u, d_u}^{-1}\Delta_k\right) \\ \times \exp\left(B'_{kx}\Delta_k + B'_{kz}d_{kz} - \frac{1}{2}B'_{kx}\Sigma_{d_u, d_u}^{-1}B_{kx}\right).$$

Then

$$\Pr\{Y_k | \Delta_k, d_{kx}, d_{kz}, (T_k = 1)\} \\ = \exp\left(B'_{kx}\Delta_k + B'_{kz}d_{kz} - \frac{1}{2}B'_{kx}\Sigma_{d_u, d_u}^{-1}B_{kx}\right) \\ \div \sum_{Y_k \text{ s.t. } T_k=1} \exp\left(B'_{kx}\Delta_k + B'_{kz}d_{kz} \right. \\ \left. - \frac{1}{2}B'_{kx}\Sigma_{d_u, d_u}^{-1}B_{kx}\right) \\ = \exp\left[\sum_{j=2}^{M+1} Y_{kj} \{\delta'_{kj}\beta_x + (z_{kj} - z_{k1})'\beta_z\} \right. \\ \left. - \frac{1}{2}\sum_{j=2}^{M+1}\sum_{j'=2}^{M+1} Y_{kj}Y_{kj'}\beta'_x\Sigma_{d_u, j, j'}\beta_x\right] \\ \div \sum_{Y_k \text{ s.t. } T_k=1} \exp\left[\sum_{j=2}^{M+1} Y_{kj} \{\delta'_{kj}\beta_x + (z_{kj} - z_{k1})'\beta_z\} \right. \\ \left. - \frac{1}{2}\sum_{j=2}^{M+1}\sum_{j'=2}^{M+1} Y_{kj}Y_{kj'}\beta'_x\Sigma_{d_u, j, j'}\beta_x\right],$$

where δ_{kj} denotes the j th $p_1 \times 1$ vector element of Δ_k . Noting that $Y_{kj}Y_{kj'} = 0$ when $j \neq j'$, $Y_{kj}^2 = Y_{kj}$, and $\Sigma_{d_u, j, j} = \Sigma_{d_u, 2, 2}$ for $j \geq 2$, the above expression reduces to

$$\left[1 + \sum_{j=2}^{M+1} \exp\left\{\left(\delta_{kj} - \frac{1}{2}\Sigma_{d_u, 2, 2}\beta_x\right)'\beta_x \right. \right. \\ \left. \left. + (z_{kj} - z_{k1})'\beta_z\right\}\right]^{-1}$$

under the usual convention that the observed case is designated as the first subject in each matched set. Recall that $\Delta_k = d_{kw} + \Sigma_{d_u, d_u} B_{kx}$ is held fixed at the value for the observed data, i.e., $Y_{k1} = 1, Y_{kj} = 0, j \geq 2$. Then $\Delta_k = d_{kw}$ and δ_{kj} reduce to $w_{kj} - w_{k1}$, which henceforth will be denoted by d_{kjwt} . Multiplying terms $\Pr\{Y_k | \Delta_k, d_{kx}, d_{kz}, (T_k = 1)\}$ over K matched sets, the full conditional likelihood is

$$\Pr \left\{ Y_1, Y_2, \dots, Y_K \mid (\Delta_k, d_{kx}, d_{kz}, T_k = 1)_{k=1}^K \right\} \\ = \prod_{k=1}^K \left\{ 1 + \sum_{j=2}^{M+1} \exp(\gamma'_{kj} \beta_x + d'_{kjt} \beta_z) \right\}^{-1},$$

where $\gamma_{kj} = d_{kjwt} - (1/2)\Sigma_{d_u, 2, 2}\beta_x$ and d_{kjt} is defined analogously to d_{kjwt} .

APPENDIX B

Numerical Solution Methods

Here we address the problem of solving for β that maximizes a likelihood

$$L(\beta) = \prod_{k=1}^K \left[1 + \sum_{j=2}^{M+1} \exp(c'_{kj} \beta) \right]^{-1}.$$

For the case of no measurement error,

$$c_{kj} = r_{kj} - r_{k1}, \quad k = 1, 2, \dots, K; j = 2, 3, \dots, M+1. \quad (B.1)$$

For the conditional scores approach, $\beta = (\beta'_x, \beta'_z)'$ and

$$c_{kj} = (\gamma'_{kj}, d'_{kjt})', \quad k = 1, 2, \dots, K; j = 2, 3, \dots, M+1. \quad (B.2)$$

By Newton's method, we iteratively solve for the maximizer of $L(\beta)$, denoted $\hat{\beta}$, using

$$\beta_{n+1} = \beta_n - [\nabla^2 g(\beta_n)]^{-1} \nabla g(\beta_n), \quad (B.3)$$

where ∇g is the gradient vector and $\nabla^2 g$ is the Hessian matrix for g . Under no measurement error, apply (B.3) directly after substituting (B.1). For the conditional scores method, we propose nesting (B.3) within a series of iterations as follows.

Step 1: With an initial guess $\beta^{(0)} = (\beta_x^{(0)'}, \beta_z^{(0)'})'$, iterate using (B.3) with $c_{kj} = (w'_{kj} - w'_{k1}, z'_{kj} - z'_{k1})'$, $k = 1, 2, \dots, K; j = 2, 3, \dots, M+1$, to obtain $\beta^{(1)} = (\beta_x^{(1)'}, \beta_z^{(1)'})'$ as though there were no measurement error in the covariates.

Step 2: Set $\gamma_{kj}^{(1)} = d_{kjwt} - (1/2)\Sigma_{d_u, 2, 2}\beta_x^{(1)}$, $k = 1, 2, \dots, K; j = 2, 3, \dots, M+1$.

Step 3: For $m = 2, 3, \dots$,

(a) Iterate using (B.3) evaluated at $\gamma_{kj} = \gamma_{kj}^{(m-1)}$, $k = 1, 2, \dots, K; j = 2, 3, \dots, M+1$ until convergence to $\hat{\beta}^{(m)}$.

(b) Update $\gamma_{kj}^{(m)} = d_{kjwt} - (1/2)\Sigma_{d_u, 2, 2}\hat{\beta}_x^{(m)}$, $k = 1, 2, \dots, K; j = 2, 3, \dots, M+1$.

(c) Repeat steps (a) and (b) until $\|\hat{\beta}^{(m+1)} - \hat{\beta}^{(m)}\| < \epsilon \|\hat{\beta}^{(m+1)}\|$.

As discussed by Stefanski and Carroll (1987), maximizing the likelihood directly for β after substituting for each γ_{kj} its expression in terms of β does not lead to the desired solution. One must solve the conditional score equations holding the $\{\gamma_{kj}\}$ fixed, and fortunately this can be done with standard conditional logistic regression software.

Because the conditional score equations are unbiased estimating equations, regularity conditions will ensure the existence of a consistent solution. However, as Stefanski and Carroll (1987) point out in a more general setting, there is not a guarantee of a unique solution, and some of the solutions may not be consistent. While there is no definitive solution to this problem, in practice, they found that, when multiple solutions are detected, the solution closest to the naive estimator (ignoring measurement) error will often be a good choice if the measurement error is not too large.