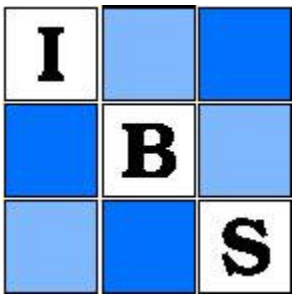

BIOMETRICS 53, 395–409
June 1997

Fully Parametric and Semi-Parametric Regression Models for Common Events with Covariate Measurement Error in Main Study/Validation Study Designs

Donna Spiegelman[1,2] and Mario Casella[1]

Departments of [1]Epidemiology and [2]Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, U.S.A.

SUMMARY

The derivation of the likelihood function for binary data from two types of main study/validation study designs where model covariates are measured with error is elaborated. Rather than limiting consideration to a restricted family of models with convenient mathematical properties, we suggest that empirical considerations, customized to the data at hand, should drive model choices. The joint likelihood function for the main study, in which the covariates are measured with error, and the validation study, in which they are not, is maximized, and estimation and inference proceeds using standard theory. Although the choice of the measurement error model is driven by empirical considerations, the relatively small validation study sizes typically seen may lead to misspecification, resulting in bias in estimation and inference about exposure–disease relationships. By using a nonparametric form for the measurement error model, the resulting semi-parametric methods suggested by Robins, Rotnitzky, and Zhao (1994, *Journal of the American Statistical Association* **89,** 864–866) and Robins, Hsieh, and Newey (1995, *Journal of the Royal Statistical Society, Series B* **57,** 409–424) are free from bias due to misspecification of the measurement error model, trading efficiency for robustness as usual. These fully and semi-parametric methods are illustrated with a detailed example from a main study/validation study of the health effects of occupational exposure to chemotherapeutics among pharmacists (Valanis et al., 1993, *American Journal of Hospital Pharmacy* **50,** 455–462). A constant prevalence ratio model for common binary events, with gamma covariate measurement error, is derived and empirically verified by the available data. A careful reanalysis of the data, taking measurement error fully into account, leads to a threefold increase in the log relative risk and no loss of statistical power. The semi-parametric estimates are consistent with the parametric results, providing reassurance that important bias due to misspecification of the measurement error model is unlikely.

1. Introduction

Covariate measurement error is a pervasive and significant source of bias in much epidemiologic research, because obtaining accurate exposure data in free-living human populations over long, or even short, periods of time is exceedingly difficult. Many methods for study design and data analysis that correct for this bias in estimation and inference for epidemiologic applications have been suggested, and recent reviews of these have been published (Armstrong, 1990; Clayton, 1991; Thomas, Stram, and Dwyer, 1993; Spiegelman, 1994). The method of maximum likelihood and the closely related method of maximum pseudo-likelihood are two of these analytic methods. Maximum pseudo-likelihood methods treat as known the parameters of a measurement error process, whose mathematical form is assumed *a priori,* or plug-in estimates of the assumed model's parameters (Gong and Samaniego, 1981). Full likelihood methods fit the form of the measurement error process to validation study data and jointly estimate the parameters of all underlying models simultaneously. Several authors have elaborated a pseudo–likelihood-based approach to estimation

Key words: Binary data; Covariate measurement error; Maximum likelihood; Semi-parametrics; Validation study.

and inference for binary data with covariate measurement error for specific parameters of interest and with a specific measurement error model (Carroll et al., 1984; Schafer, 1993; Whittemore and Grosser, 1986). Richardson and Gilks (1993a, 1993b) applied conditional independence models arising in a Bayesian perspective with Gibbs sampling to focus primarily on parameter estimation from the full likelihood.

Because validation studies are typically small (e.g., no more than 10% of the total sample), there is concern that the measurement error process will be misspecified, introducing bias in estimation and inference. Semi-parametric likelihoods and estimating equations can be constructed that are robust to misspecification of the model for the measurement error process. Carroll and Wand (1990) used a kernel density estimate of the measurement error model and Flanders and Greenland (1991) proposed a Horvitz–Thompson type estimator (Horvitz and Thompson, 1952) in the validation study alone. More recently, Robins, Rotnitzky, and Zhao (1994) and Robins, Hsieh, and Newey (1995) proposed semi-parametric estimators appropriate for this setting that, depending on the exact form of the estimating equations, may be semi-parametric efficient for all possible measurement error models, in the sense of Begun et al. (1983), or locally semi-parametric efficient.

In this paper we focus on obtaining both point estimates of model parameters uncontaminated by bias due to covariate measurement error and efficient, correctly centered confidence intervals with the specified coverage probability. We present fully parametric likelihood-based estimation and inference with covariate measurement error and common binary events using data from main study/validation study designs. We compare results from the Horvitz–Thompson estimator and the semi-parametric locally efficient estimator to results obtained from the fully parametric likelihood. The methodology is formulated for main study/validation study designs, where a random subsample of the data is available for modeling the covariate measurement error process. The validation data make it possible both to estimate consistently the parameters of interest without invoking empirically unverifiable assumptions about the measurement error process and to represent correctly the true uncertainty in the resulting inference using confidence intervals that reflect all sources of variability in the data, including those due to covariate measurement error. The usefulness of main study/validation designs hinges upon the existence of a 'gold standard' for assessing exposure, which can be feasibly measured at least in a small sample.

In Section 2.2 the methodology for fully parametric maximum likelihood methods for a main study/validation study of a binary outcome is elaborated. In Section 2.3 semi-parametric methods are reviewed, and simplified computational formulas are presented that are appropriate for the binary outcome data setting. Section 3 develops a detailed example from occupational epidemiology, with gamma covariate measurement error. Section 4 contains a brief discussion, highlighting directions for future research.

2. Methods

2.1 *The Main Study/Validation Study Design*

Consider the usual study in epidemiologic research. There are n_1 subjects with data on a binary outcome variable D, where 1 indicates the presence of an event of interest and 0 otherwise. The occurrence of D depends on the values of two covariate vectors, x and U_1, where $\dim(x) = p$, $\dim(U_1) = q$, and x is perfectly measured in the validation study, but observed with the error in the main study, while U is always perfectly measured. There is also a set of covariates U_2 on which the measurement error processes may depend, but which are conditionally independent of the occurrence of D given (x, U_1). Some of the covariates in U_2 may be the same as those in U_1, and the unique elements of U_1 and U_2 together are denoted by U. In the main study, these n_1 subjects are observed with data (D_i, X_i, U_i), $i = 1, \ldots, n_1$, where $\dim(X_i) = p$ and X_i correspond on an element-wise basis to the covariates (x_i) observed with error. To complete this main study/validation study design, a subsample of n_2 additional subjects is available with one of two data structures: (D_i, x_i, U_i, X_i), $i = n_1+1, \ldots, n_1+n_2$, or (x_i, U_{2i}, X_i), $i = n_1+1, \ldots, n_1+n_2$. The first structure describes an internal validation study, while the second describes an external validation study (Carroll and Stefanski, 1990). The internal validation study is an augmented simple random sample from the main study. The external validation study may originate from another source, in which case an empirically unverifiable assumption is invoked—that the measurement error model that underlies the measurement error process observed in the validation data is the same as the unobservable one that generated the main study data.

2.2. *Specification of the Relative Risk Model*

The investigator must propose a model for $\Pr(D \mid x, U_1; \beta) = f_1(D \mid x, U_1; \beta)$. When x is observed without error in the $n_1 + n_2$ study subjects, this model would be used to estimate the relative risk,

defined as the ratio of the estimated probabilities of occurrence of D corresponding to a Δ unit increase in value for the corresponding model covariate. When D is sufficiently rare, the relative risk can be well approximated by the odds ratio, making the logistic regression model a convenient choice in many epidemiologic studies. When the frequency of occurrence of D is high, the odds ratio, itself a parameter of no intrinsic scientific interest (Greenland, 1987), is no longer a good approximation to the prevalence ratio or risk ratio. The constant prevalence (risk) ratio model must then be directly fit using form

$$f_1(D = 1 \mid x, U_1) = e^{\beta_0 + \beta_x' x + \beta_{U_1}' U_1}. \tag{1}$$

If x were observed without error, the prevalence ratio would be estimated by exponentiating $\hat{\beta}_x \Delta$, where Δ is an appropriate increment. Model (1) has the restriction on the allowable parameter space that $\beta = (\beta_0, \beta_x, \beta_{U1})$ satisfy $\Pr(D = 1 \mid x, U_1) \leq 1$. In epidemiology, the logistic and constant relative risk models are nearly universally assumed by biomedical investigators, and empirical verification of fit as compared with other possible forms is uncommon. The fully parametric or semi-parametric methods considered in this paper require that this model be parametrically specified.

2.3 *Fully Parametric Maximum Likelihood Methods for Estimation and Inference*

2.3.1 *Specification of the measurement error model.* The next step is to fit an appropriate measurement error model, f_2, to the data. First, we treat the case of nondifferential measurement error, i.e., where $f(X \mid x, U_2, D; \theta) = f(X \mid x, U_2; \theta)$, or equivalently, where $f_1(D \mid x, X, U_1, U_2; \beta) = f_1(D \mid x, U_1; \beta)$. In Section 2.3.2 we will discuss empirical verification of this assumption and methods that are appropriate when this assumption is not verified. In the validation study, measurement error models $f_2(x \mid D, X, U_2; \theta)$ are successively proposed and fit to the data until one that is considered adequate is found. Usually, there is no intrinsic interest in f_2; its form and its parameters are nuisances.

The process of specifying f_2 involves identification of both the family of models to which f_2 belongs, as well as the functional form of its parameters within this family. Comparisons of the values of the log-likelihoods with several choices for f_2 can be used to select the best model for the data at hand from among its competitors, in combination with criteria derived from graphical and other methods to judge goodness-of-fit. Take, for example, the case where $\dim(x) = 1$. We may successively fit models from the normal ($\theta = \mu, \sigma^2$), log-normal ($\theta = \mu, \sigma^2$), or gamma ($\theta = \mu, \nu$) families to the validation data. Preliminary analysis will then indicate which one is best. We may then proceed to fit the mean parameter μ as a linear function of the covariates and use a liberal forward variable selection procedure to identify which components of (X, U) are the strongest determinants of the mean. Identifying the appropriate measurement error model is a challenging data analysis task in which all the thorny issues of data analysis arise.

2.3.2 *Derivation of the model for $D \mid X, U$.* If measurement error is nondifferential, given f_1 and f_2, $f_3(D \mid X, U; \beta, \theta)$ is derived as

$$f_3(D \mid X, U; \beta, \theta) = \int f_1(D \mid x, U_1; \beta) f_2(x \mid X, U_2; \theta)\, dx, \tag{2}$$

where integration is over the range of x and may, of course, be multidimensional. Even if f_1 depends only on U_1, f_3 will depend on U, the unique elements of U_1 and U_2. The model f_3 may not be integrable in closed form and thus may pose challenging numerical problems. When f_1 is logistic and f_2 is normal, see Crouch and Spiegelman (1990) for a fast numerical solution.

If measurement error is differential, f_3 is derived as follows:

$$f_3(D \mid X, U; \beta, \theta, \gamma) = \int f_1(D \mid X, U_1; \beta) f_2(X \mid D, x, U_2; \theta) f_4(x; \gamma)\, dx. \tag{3}$$

Now, f_2 must be specified jointly conditional upon (x, U_2, D), and the marginal density of $f_4(x; \gamma)$ must additionally be specified. In prospective studies, the possibility of the dependency of the measurement error model on D is often eliminated by design; in cross-sectional studies, f_2 may depend on D as well as (X, U_2). To investigate the empirical evidence for differential measurement error, standard statistical procedures can be used. Suppose it is reasonable to hypothesize that f_2 depends on D through the parameter ϕ. Then, the hypothesis H_0: $\phi = 0$ can be tested vs. its alternative, H_A: $\phi \neq 0$, using either the full data likelihood (see Section 2.2.3 for its construction) or the likelihood from the validation study alone, the latter being computationally simpler, the former, more efficient. An example of this is given in Section 3.3. If differential measurement error

is indicated, than a model for $f_4(x;\gamma)$ must be fit to the marginal distribution of x in the validation study, in addition to the conditional model for $X \mid D, x, U_2$. Note that, in these circumstances, conditioning of (x, X) in f_2 is reversed.

Because f_3 contains both β and θ, the main study will also contain information about the measurement error model. This has implications for the overall modeling strategy. In the example given in Section 3, we calculate the increase in asymptotic relative efficiency (ARE) in the MLEs for the measurement error model parameters from using the $n_1 + n_2$ main study/validation study subjects, compared to using the n_2 validation study subjects alone. To the extent that the increase in ARE is nontrivial, an iterative model strategy is implied, where, for convenience, model building for f_2 is first conducted in the validation study alone. Once an adequate model is found, this model is then fit to the full data set and further adjustments made as necessary. This strategy is particularly convenient when f_2 is supported by standard statistical software, automating the generation of diagnostics, variable selection, and computation of other statistical quantities of interest. Since normal, gamma, and log-normal families are natural choices for measurement error models of continuous data, it will often be the case that standard software can be utilized at this stage.

It will rarely be the case that likelihood analysis using f_3 will be supported by standard software.

2.3.3 *Estimation and inference.* Once f_1 is specified, an appropriate form for f_2 is found through preliminary analysis using the validation data, and f_3 is derived, full likelihood-based parameter estimation and inference can proceed. In the main study/internal validation study design, this involves maximizing

$$L_{IV}(\beta,\theta) = \sum_{i=1}^{n_1} \log[f_3(D_i \mid X_i, U_i; \beta, \theta)] + \sum_{i=n_1+1}^{n_1+n_2} \log[f_2(x_i \mid X_i, U_{2i}; \theta)] + \sum_{i=n_1+1}^{n_1+n_2} \log[f_1(D_i \mid x_i, U_{1i}; \beta)] \tag{4}$$

to obtain $(\hat{\beta}_{ML}, \hat{\theta})$. For the main study/external validation study design, the likelihood function is of the form

$$L_{IV}(\beta,\theta) = \sum_{i=1}^{n_1} \log[f_3(D_i \mid X_i, U_i; \beta, \theta)] + \sum_{i=n_1+1}^{n_1+n_2} \log[f_2(x_i \mid X_i, U_{2i}; \theta)].$$

The maximum likelihood estimators are consistent estimators of the true values of the underlying parameters, with a variance–covariance matrix that can be estimated as the inverse of the observed information matrix. Wald-type confidence intervals can be formed for inference or profile likelihood-based confidence intervals can be calculated. The two-sided Wald-type hypothesis test statistic can be used for testing the hypothesis H_0: $\beta_{x_r} = 0$ with the variance of $\hat{\beta}_{x_r}$ estimated in one of the several ways suggested by standard theory, where r indexes one element of the p-dimensional vector x. Alternatively, a likelihood ratio test procedure can be used.

2.4 *Semi-Parametric Methods for Estimation and Inference*

Semi-parametric methods avoid any specification for f_2. For $\hat{\beta}_{SP}$, the Horvitz–Thompson version of this approach, a weighted likelihood function based on $f_1(D \mid x, U_1; \beta)$ is maximized to obtain $\hat{\beta}_{SP}$ using the validation data alone, with weights equal to the inverse of the empirical selection model, $\Pr(V_i = 1 \mid D_i, X_i, U_i) = \pi_i$, for each subject i, where $V = 1$ indicates selection into the validation study. When one or more components of (D, X, U) are continuous, these selection probabilities can be estimated from a selection model fit to the $n_1 + n_2$ study subjects (Robins et al., 1994). A robust variance estimator is used to quantify the uncertainty around $\hat{\beta}_{SP}$. Because validation studies are explicitly designed and controlled by the investigator and sampling is typically completely at random, $\pi_i = \Pr(V_i = 1 \mid D_i, X_i, U_i) = n_2/(n_1 + n_2)$. With completely random sampling, the Horvitz–Thompson estimator $\hat{\beta}_{SP}$ simply maximizes

$$\begin{aligned} L_{SP} &= \sum_{i=n_1+1}^{n_1+n_2} \pi_i^{-1} \log[f_1(D_i \mid x_i, U_{1i}; \beta)] \\ &= \pi^{-1} \sum_{i=n_1+1}^{n_1+n_2} \log[f_1(D_i \mid x_i, U_{1i}; \beta)], \end{aligned} \tag{5}$$

which is equivalent to maximizing the likelihood based upon f_1 in the validation study. This estimator avoids specifying f_2 altogether. In studies of rare events, this estimator may be infeasible because there will be a small number of events, perhaps none, in the validation study. Even when the event is common, this estimator may be inefficient since validation studies are typically small (e.g., no more than 10% of the overall study population).

Robins et al. (1995) proposed an estimator that is consistent and semi-parametric efficient (Begun et al., 1983), i.e., it is the minimum variance, consistent estimator over all possible models for f_2. This approach is semi-parametric in exactly the sense needed for the measurement error setting, i.e., where f_1 is parametric and f_2 is nonparametric. Because f_2 is almost always unknown *a priori,* methods that are robust to misspecification of f_2 are ideal. Although their solution applies widely to models of continuous and categorical outcomes, to nonrandom validation study sampling, and to mismeasured covariates of arbitrary dimension, in the case of a binary outcome with random validation study sampling and a single mismeasured covariate, their formulas simplify considerably (Appendix 1). The estimator $\hat{\beta}_{SPLE}$, given in Appendix 1, is the semi-parametric locally efficient form of their class of estimators. This estimator consists of two pieces, $\mathbf{U}_1$ and $\mathbf{U}_2$, which sum to a quantity of expectation 0. The first piece, $\mathbf{U}_1$, is the score equations from the likelihood of validation study for f_1, and, under random validation study sampling, is identical to the equations for which $\hat{\beta}_{SP}$ is the solution. The second piece uses additional data from both the main study and the validation study. Any choice of the function ϕ leads to consistent estimates of β in (A1.1); Robins et al. (1995) found forms of ϕ that lead to estimators for β that are semi-parametric efficient and semi-parametric locally efficient. To avoid the computational complexity and intensity required for semi-parametric efficiency, that is, to avoid the need for nonparametric regression techniques to estimate f_2, a parametric version of f_2 can be used instead to obtain $\hat{\beta}_{SPLE}$. When the parametric version of f_2 used is close to the true underlying f_2, $\hat{\beta}_{SPLE}$ will be nearly semi-parametric efficient. Even when f_2 is misspecified, $\hat{\beta}_{SPLE}$, unlike $\hat{\beta}_{ML}$, is consistent.

3. Example: Acute Health Effects of Occupational Exposure to Chemotherapeutic Agents among Pharmacists

In a study of acute effects of occupational exposure to chemotherapeutic agents among 675 pharmacists and pharmacy technicians (Valanis et al., 1993), 56 participants (8%) were randomly subsampled in an internal validation study of the exposure variable. The exposure of interest is the usual number of doses of chemotherapeutic agents mixed per week. In the main study, where $n_1 = 675 - 56 = 619$, X was self-reported on questionnaires, and in the validation study, where $n_2 = 56$, x was assessed through a written diary that was maintained concurrently with work activity for 1–2 weeks duration. Although data were available on 27 acute health symptoms, in this example we examined the effect of exposure on the prevalence of fever in the past 3 months. There were several possible additional independent risk factors (U) for this outcome (D), including sex, age, body mass index, occupational stress level, type of workplace, type of work shift, and current cigarette-smoking status. Data on these variables were available for each study participant from self-reported questionnaires. In this example, as in the original analysis, it was assumed that these covariates are measured without error.

3.1 *Specification of the Relative Risk Model* $f_1(D \mid X, U_1; \beta)$ *and Identification of Model Covariates* U_1

The prevalence ratio was the parameter of interest here, and it was assumed that the form of f_1 was such that the prevalence ratio was constant over all levels of the other determinants of D. The prevalence of fever was high enough overall (16%) that we did not wish to use the odds ratio approximation to the prevalence ratio, so we used the constant prevalence ratio (CPR) model (1) for f_1. Once the full likelihood was developed as described subsequently in Sections 3.2 and 3.3, we addressed the issue of variable selection for U_1. Using a forward selection procedure based on the likelihood ratio test criterion, we identified the appropriate components of U_1 using the likelihood function (10). Three variables were clearly and independently associated with fever: age, working in a community hospital vs. other locations, and working day shift only, and these variables comprise U_1 in all further analysis.

3.2 *Fitting of the Measurement Error Model* f_2

There was evidence for moderate exposure measurement error in these data, and the simple correlation between x and X estimated in the validation study was 0.70 (Figure 1). We used

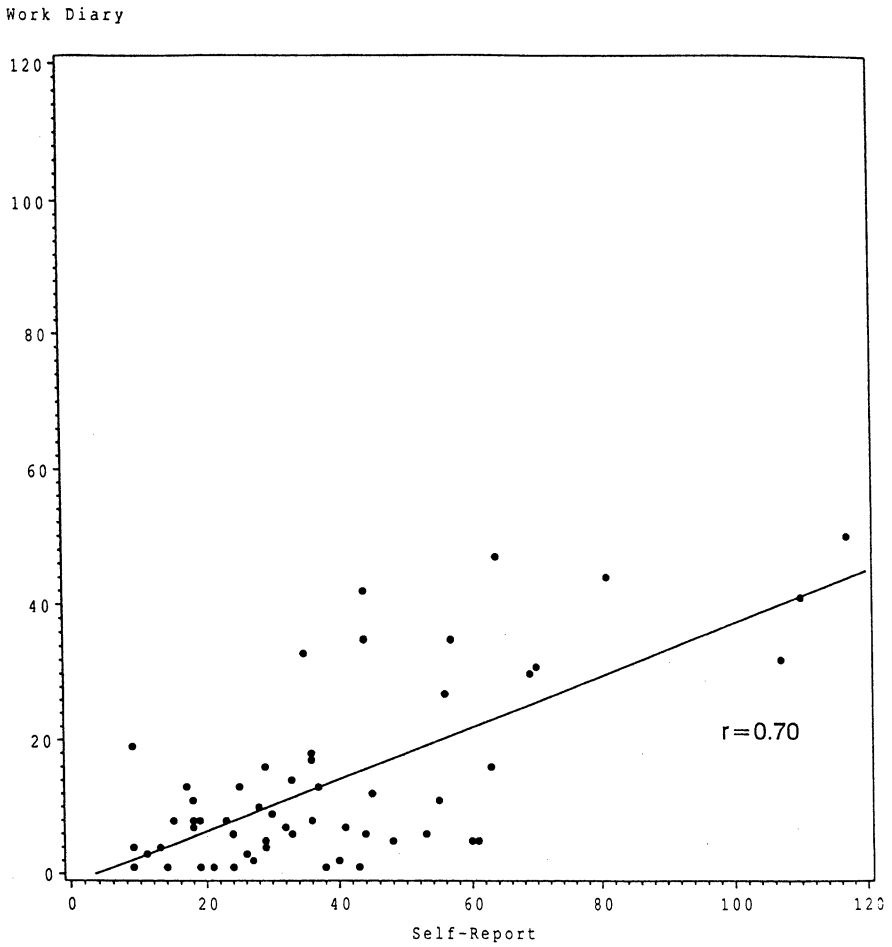

Figure 1. Number of drugs mixed per week, work diary vs. self-report ($n_2 = 56$).

the family of gamma distributions to model f_2, i.e.,

$$f_2(x \mid X, U_2) = \Gamma[\mu(X, U_{2,\mu}), \nu(X, U_{2,\nu})] = \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu}\right)^{\nu} x^{\nu-1} e^{-\frac{\nu}{\mu}x}. \tag{6}$$

Unlike, for example, the corresponding normal distribution, this conditional distribution for $x \mid (X, U_2)$ can be fit so as not to produce negative values of x for X at or near 0, even if the variance is large, and, unlike the log-normal, this family of distributions is well defined at $x = 0$. In addition, if a term for X is included in the specification of μ and/or ν, this model will incorporate a dependency of the error variance on the level. As shown in Figure 2, the marginal distributions of X in the complete data set $(n_1 + n_2)$ and x in the validation study were sharply skewed, and the data are concentrated near 0.

We fit gamma distributions to the data with several specifications for $\mu(X, U_{2\mu})$ and $\nu(X, U_{2\nu})$. We first did this in the validation study alone, although, as discussed in Section 2.3.1, there was at least some information about the specification of f_2 in the main study data as well. However, we will show shortly that, in this example, the amount of information about f_2 in the main study was small. We used likelihood ratio testing procedures to choose a model from the class $\mu(X, U_{2\mu}) = a + bX^c$ and $\nu(X, U_{2\nu}) = d + eX^f$. When $a = 0$ and $b = 1$, a multiplicative model for μ is obtained, and when $c = 1$, an additive model for μ results. Both in the validation study alone and in the full data set, using the likelihood ratio test criterion, a reasonable fit to the data was obtained when μ and ν were both additive in X. Although the multiplicative model for μ fit the data approximately as well as the additive one, this model led to difficult numerical problems of several kinds. Because the estimated value of β_x was insensitive to the form of μ, we preferred the additive model for computational convenience.

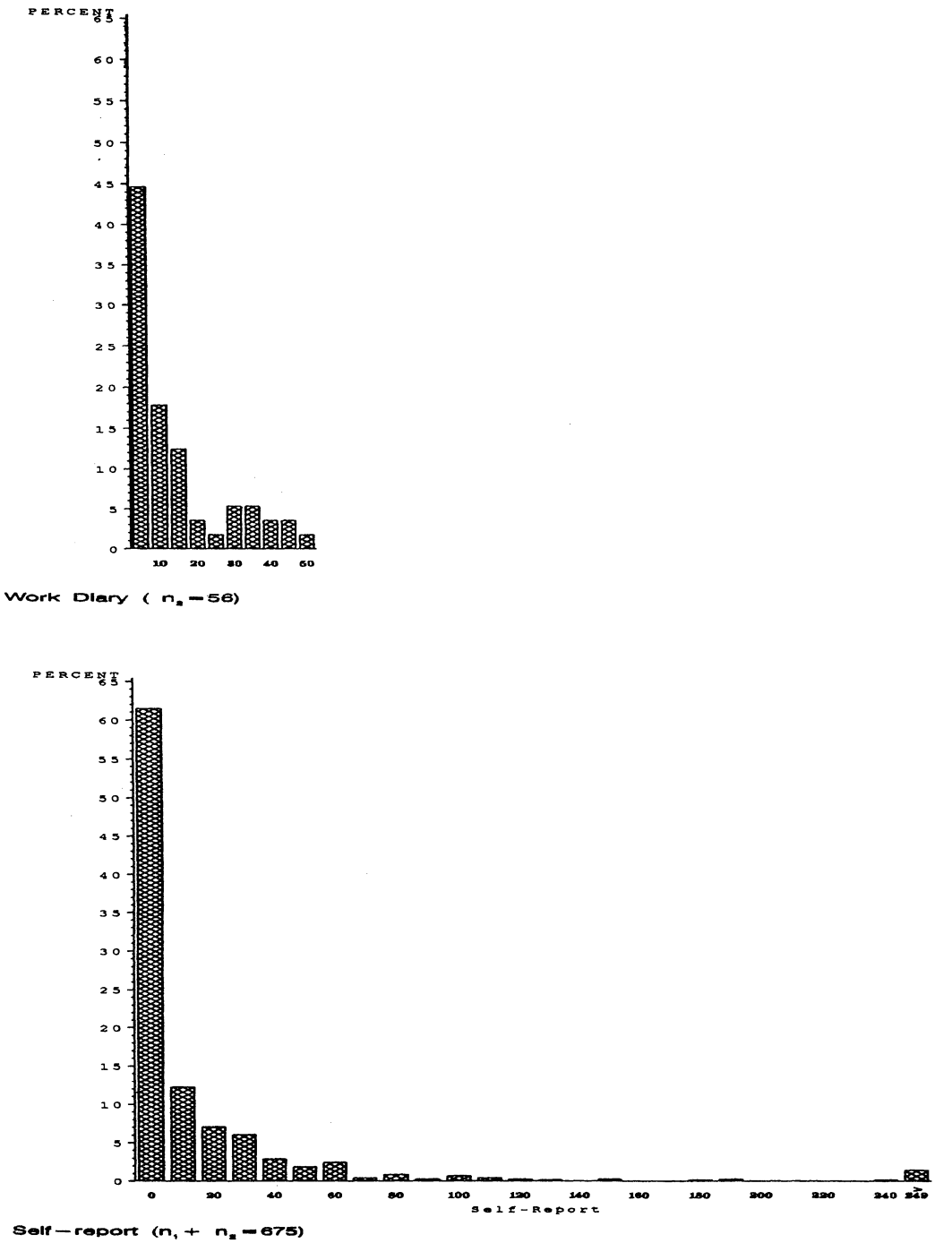

Figure 2. Distribution of drugs mixed per week.

Next, we needed to select from among the available covariates U the ones that should be included in $U_{2\mu}$ and $U_{2\nu}$. We used a forward variable selection procedure, first to find the variables to be included in $U_{2\mu}$, where

$$\mu = a + b_X X + b'_{U_{2\mu}} U_{2\mu} \tag{7}$$

and $\nu = d + eX$. Given the model selected for μ, we then proceeded to fit the model for ν, where

$$\nu = d + e_X X + e'_{U_{2\nu}} U_{2\nu}, \tag{8}$$

in the same forward selection fashion. We found that conditional on first including X, after including age, current smoking status, educational status, type of shift (day vs. night, rotating) in $U_{2\mu}$, no additional variable improved the fit of the data to the model for μ. There were no variables that improved the fit of the data to the model for ν (i.e., $U_{2\nu}$ was empty).

We produced plots of the standardized deviance residuals of the models fit to the data using (10) for the n_2 validation study subjects, using the standardized gamma deviances (Cox and Snell, 1968). We used the plot of the standardized deviance residuals vs. the predicted x's, $\hat{x}_i$, to check for systematic departures from the fit of the link function (McCullagh and Nelder, 1983, Section 11.5.2) and the plot of the absolute value of the standardized deviance residuals vs. $\hat{x}_i$ to check for systematic departures from the fit of the variance function (McCullagh and Nelder, 1983, Section 11.5.3). In neither case was any pattern visible to indicate systematic departures from the selected model form.

3.3 *Derivation of f_3 and Maximization of the Likelihood*

Next, we needed to derive an expression for $f_3(D \mid X, U)$. To use (2) rather than the more complex (3) to derive f_3, we needed to verify the conditional independence assumption that $f_1(D \mid x, X, U_1) = f_1(D \mid x, U_1)$. Using the validition study data, we compared the fit of the model $f_1(D = 1 \mid x, X, U_1) = e^{\beta_0+\beta_1 x+\beta_2 X+\beta'_{U_1}U_1}$ to the model $f_1(D = 1 \mid x, U_1) = e^{\beta_0+\beta_1 x+\beta'_{U_1}U_1}$, where U_1 was selected as described in Section 3.1, to test the hypothesis H_0: $\beta_2 = 0$. Using the likelihood ratio test, the p-value was 0.36. Having thus empirically verified the conditional independence assumption, we proceeded to derive f_3 using (2), given f_1 and f_2, as

$$f_3(D = 1 \mid X, U) = e^{\beta_0+\beta'_{U_1}U_1}\left[\frac{\nu}{\nu - \mu\beta_x}\right]^{\nu}, \tag{9}$$

where μ and ν are given by (7) and (8), respectively. The derivation for f_3 is given in Appendix 2. The specifications of f_1–f_3 thus identified, we then constructed the likelihood:

$$\begin{aligned} L_{ACE}[\beta, \mu(a,b), \nu(d,e)] = &\sum_{i=1}^{n_1} \log\{[f_3[D_i \mid X_i, U_i; \beta, \mu(a,b), \nu(d,e)]]\} \\ &+ \sum_{i=n_1+1}^{n_1+n_2} \log\{[f_2[x_i \mid X_i, U_{2i}; \mu(a,b), \nu(d,e)]]\} \\ &+ \sum_{i=n_1+1}^{n_1+n_2} \log[f_1(D_i \mid x_i, U_{1i}; \beta)], \end{aligned} \tag{10}$$

where f_1, f_2, and f_3 are given by equations (1), (6–8), and (9) respectively, and U_1 = (age, community hospital, shift), $U_{2\mu}$ = (age, current smoking status, educational status, shift) and $U_{2\nu}$ = (). L_{ACE} is thus a 13-parameter model, where $\dim[\beta = (\beta_0, \beta_x, \beta_{U1})] = 5$, $\dim[\mu = a + b'(X,$ age, current smoking status, educational status, shift)$] = 6$, and $\dim[\nu = d + eX] = 2$. There are five parametric constraints on the model implied by likelihood function (10) that must be considered in the fitting procedure (Appendix 3). The feasible sequential quadratic programming algorithm (FSQP) developed by Zhou and Tits (Panier and Tits, 1993) constrain all trial values of the parameters to remain within the feasible region.[1] For likelihood function (10), this feature was critical because values outside of the bounds of the constrained parameter space will lead to undefined mathematical operations in the real number space. The FSQP algorithm worked best when a subroutine for the gradient of the objective function to be maximized was supplied, although this was not required. Wald-type inference requires the Hessian of the objective function. Rather than going through the tedious and error-prone steps of algebraically deriving and then computer coding the gradient and Hessian of this complicated 13-parameter likelihood function (10), we exploited the automatic differentiation technology developed by Bischof and colleagues[2] (Bischof et al., 1992) to produce computer-generated gradient and Hessian subroutine code for (10). The ML analysis converged to a point in the interior of the parameter space, i.e., with all constraints satisfied, using these algorithms.

[1] A user-friendly shareware routine which implements these features is available over the Internet from Dr. André Tits, `andre@src.umd.edu`

[2] To obtain access to this software, address correspondence to Dr. Christian Bischof, `bischof@mcs.anl.gov`

3.4 *Calculation of $\hat{\beta}_{SPLE}$, the Semi-Parametric Locally Efficient Estimator*

The semi-parametric locally efficient estimator and related quantities were obtained using a subroutine[3] that computes the estimating equations to which this estimator is the solution (Appendix 1), for arbitrary f_1, f_2 independent of D, arbitrary validation selection model (π_i), $i = 1, \ldots, n$, D binary, and $\dim(x) = 1$. The variance of the semi-parametric locally efficient estimator is estimated by $I^{-1} \Sigma_{i=1}^{n_1+n_2} [U_i(\beta)U_i(\beta)']I^{-1} \mid_{\hat{\beta}_{SPLE}}$, where $I = \partial U(\beta)/\partial\beta$ and $U(\beta)$ is given by (A1.1). FSQP was again used to find this estimator by minimizing the objective function $g(\beta) = U(\beta)'U(\beta)$. The SPLE analysis converged in 86 seconds of CPU time on a SPARC IPX workstation with a Weitek processor, using $\hat{\beta}_{ML}$ as initial values.

3.5 *Results*

The results of fitting this model and calculating key inferential quantities with respect to the quantity of primary scientific interest in this study, β_x, are given in Table 1. In this table we compared results from four procedures. In the uncorrected (UC) analysis, we reproduced the usual analysis of n_1+n_2 participants as it would have been conducted if no validation data were available, or if the data were available but the theoretical and computational technology to do the proper analysis were not, as was the case with Valanis et al.'s (1993) original publication. The likelihood function maximized for this analysis had the form

$$L_{UC}(\tilde{\beta}) = \sum_{i=1}^{n_1+n_2} D_i \log\left[e^{\tilde{\beta}_0+\tilde{\beta}_X X_i+\tilde{\beta}'_{U_1} U_{1i}}\right] + (1 - D_i)\log\left[1 - e^{\tilde{\beta}_0+\tilde{\beta}_X X_i+\tilde{\beta}'_{U_1} U_{1i}}\right]$$

and corresponds to the uncorrected constant prevalence ratio model. The maximum likelihood (ML) analysis was based on likelihood function (10), using the full data set as described in Sections 3.1–3.3. The third analysis calculated $\hat{\beta}_{SP}$, the Horvitz–Thompson estimator, and the fourth analysis computed $\hat{\beta}_{SPLE}$, the semi-parametric locally efficient method described by Robins et al. (1995) (SPLE). The two semi-parametric analyses assumed validation study sampling was completely at random.

There was a nearly threefold increase in the log-relative prevalence ratio when the ML approach was used to correct the estimate of effect for measurement error, compared with the UC analysis, and the p-value was statistically significant. We used the interdecile range (IDR) as the increment for which the prevalence ratio was calculated since extreme quantile contrasts are the typical increment choice for continuous variables in epidemiology. The ML and SPLE estimates did not differ significantly from one another. At the value of the MLE, the SPLE was less than one tenth as efficient as the MLE. This is the price we paid for robustness against misspecification of f_2. To check that validation study selection was completely at random, as planned, we modeled the selection probabilities as a function of the data available for all $n_1 + n_2$ study subjects, and recalculated $\hat{\beta}_{SPLE}$. Empirical verification of simple random sampling into the validation study was obtained since the estimates calculated assuming simple random sampling were virtually identical to the estimates calculated without this assumption (data not shown).

Here, the SP estimate was so inefficient as to be useless. There were only six cases in the validation study, and five parameters needed to be estimated. Asymptotic results are not applicable for this estimator in this example.

3.6 *Distribution of Information Between the Main Study and Validation Study*

It was of interest to determine how much information about β is derived from the validation study. Although, in this example, the validation study is 8% of the overall sample, each validation study observation should be more informative than its main study counterpart because the exposure was measured without error. In addition, the main study contains information about the parameters of the measurement error model, and it is of interest to quantify this in some way.

Table 2 presents the estimated asymptotic relative efficiency ($\widehat{ARE}$) for the estimates of each parameter in the model used by the appropriate piece of the likelihood function (10) from the validation study alone (VS) compared with the ML analysis, where

$$\widehat{ARE}(\omega) = \frac{\widehat{\text{var}}_{n_2}^{-1}(\hat{\omega})}{\widehat{\text{var}}_{n_1+n_2}^{-1}(\hat{\omega})} \mid \Omega = \hat{\Omega}_{ML}.$$

[3] Address requests for this program to `stdls@gauss.bwh.harvard.edu`

Table 1
A comparison of four approaches to estimation and inference for the effect of an exposure measured with error

Approach	n	$\hat{\beta} \times 10^{-3}$	$\mathrm{SE}(\hat{\beta}) \times 10^{-3}$	Δ	Prevalence ratio (PR) in Δ	95% Confidence intervals for PR in Δ		p-value	
						Wald	Profile likelihood	Wald	Likelihood ratio test
UC	675	1.64	0.99	IDR[a] $(X) = 52$	1.06	0.99–1.13	1.03–1.07	0.10	0.026
ML	675	4.64	2.74	IDR $(x) = 34$	1.17	0.97–1.41	1.04–1.26	0.09	0.023
SP	56	31.69	26.49	IDR $(x) = 34$	2.94	0.50–17.2	n/a	0.23	n/a
SPLE	675	2.81	11.54	IDR $(x) = 34$	1.10	0.46–2.37	n/a	0.81	n/a

[a] IDR = interdecile range.

Table 2
Relative information content of the main study and validation study in the Valanis et al. (1993) data

Parameter[1]	$\widehat{ARE}$ (%)
β_X (exposure)	1
β_{U11} (age)	6
β_{U12} (community hospital)	4
β_{U13} (day shift)	9
b_X (exposure)	93
$b_{U21\mu}$ (current smoker)	99
$b_{U22\mu}$ (educational level)	98
$b_{U23\mu}$ (day shift)	98
$b_{U24\mu}$ (age)	100
e_X (exposure)	0

[1] β's are from $f_1(D = 1 \mid x, U_1) = \exp\{\beta_0 + \beta_x x + \beta_{U11} AGE + \beta_{U12} Community\_hospital + \beta_{U13} Day\_shift\}$. b's are from $f_s(x \mid X, U_2)$, where $x_i \sim \Gamma(\mu_i, \nu_i)$, $\mu = a_0 + b_X X + b_{U2} \cdot Current\_smoker + b_{U22} Education + b_{U23} \cdot Day\_shift + b_{U24} Age$, and $\nu = d + eX_X$.

$\widehat{\text{var}}_{n_2}(\hat{\omega})$ is the variance of $\hat{\omega}$ from the validation study alone, as indicated by the subscript n_2, and $\Omega = (\omega_r, r = 1, \ldots, 13) = (\beta, \mu, \nu)$ as specified by models f_1 and f_2 given in Sections 3.1 and 3.2. Most of the information about the parameters from the measurement error model $f_2(\mu, \nu)$ was derived from the validation study, although it should be noted that the contribution from the main study was not negligible. Because only 6 of the 104 reported cases of fever in this study occurred among validation study participants, the validation study contributed little information on β.

4. Discussion

Many investigators have pointed to computational barriers to the widespread adoption of likelihood-based methods that explicitly correct for bias due to covariate measurement error (e.g., Carroll and Stefanski, 1990; Whittemore and Keller, 1988). Gibbs sampling (Richardson and Gilks, 1993a, 1993b) and the E-M algorithm (Whittemore and Grosser, 1986) tend to be computationally inefficient methods for likelihood maximization, using many more iterations to achieve convergence at the same level of precision than the corresponding Newton algorithm would require. Since the likelihood functions used to estimate the parameters of interest must be customized to the setting from which the data arise, it is not possible to standardize software in the usual way. The availability of automatic differentiation greatly enhanced our ability to quickly find the maximum of these functions with a great deal of flexibility (Bischof et al., 1992) using the Newton method and related algorithms.

An interesting feature of the results presented in Table 1 is the large discrepancy between inferential quantities obtained from the Wald approach compared with the likelihood ratio testing and profile likelihood approach (LR). This phenomenon has been noted and studied theoretically and through simulation by other investigators. The overall conclusion from these publications is that Wald-type inference is often unreliable for complex and nonstandard models at typical sample sizes, and LR methods should be used instead (Vaeth, 1985; Moolgavkar and Venzon, 1987). Because the information content of a main study observation with respect to the parameters of interest can be so much smaller than that of a validation study observation, the typical main study/validation study design may be far from the effective sample size where large sample asymptotics dominate, even when n_1 is very large.

Because exposure assessment methodologies vary widely from exposure to exposure, as well as from study to study for a given exposure, models for the measurement error process are best developed on a case by case basis. In addition, the parameter of interest will vary. In epidemiology, common parameters of interest are the prevalence ratio, risk ratio, and hazard ratio. These features motivate the need for a flexible but computationally and conceptually straightforward methodology,

in which measurement error models of arbitrary complexity can be fit to the data at hand using all of the technologies that are available to the adept data analyst, with the final goal of producing efficient unbiased estimates of the parameter(s) of interest.

Semi-parametric modeling methodologies are currently quite popular in statistics. These methods aim to relax some of the distributional assumptions required by a fully parametric approach. The semi-parametric methods given by Robins et al. (1994, 1995) are nonparametric exactly as needed, in the specification of f_2. The efficiency cost for this robustness may be intolerably high for many applications, even when the semi-parametric locally efficient or semi-parametric efficient estimators are used. At the very least, however, these methods can be used as a diagnostic. When an inconsistency between the ML and SPLE estimates is evident, misspecification of f_2 in the ML version is a likely explanation. Because all aspects of the mathematical form of f_2 can and should be empirically verified in the main study/validation study data at hand, gross misspecifications of f_2 are, in general, unlikely. Further research on the effects of "small" departures from correct specification of f_2 in relation to the magnitude of bias in ML estimates and inference is of interest.

Both the fully parametric and semi-parametric methods discussed in this paper assume that x is perfectly measured, an assumption that is rarely met completely in practice. A recent paper suggests that certain typical deviations from this assumption may lead to little, if any, loss of consistency, but will lead to loss of efficiency (Wacholder, Armstrong, and Hartge, 1993). Further research on this issue is important to pursue. The methods discussed in this paper assume that measurement error is nondifferential, and validation study selection is completely at random. Both of these assumptions can and should be empirically verified. The semi-parametric methods developed by Robins et al. (1994, 1995) require neither of these assumptions and little additional complexity is introduced by relaxing either. As shown by (3), if nondifferential measurement error cannot be assumed, using fully parametric methods in the derivation of f_3 requires parametric identification of the marginal distribution of x in addition to the conditional distribution of x given (D, X, U_2). The efficiency advantage of the fully parametric methods will certainly be compromised in this setting, and the potential for significant misspecification of f_2, f_4, and, consequently, f_3 will be increased, possibly more seriously compromising the utility of fully parametric methods.

In conclusion, a careful analysis taking measurement error fully into account can lead to quite a different interpretation of the data. In the example, fully parametric likelihood-based measurement error correction resulted in a threefold increase in the estimated parameter of interest, the log relative risk, compared to the uncorrected analysis, and sufficient statistical power was retained. Maximum likelihood and semi-parametric locally efficient methods are shown to be useful and feasible, making a minimum of empirically unverifiable assumptions.

Résumé

Nous présentons la manière d'obtenir la fonction de vraisemblance pour des données binaires dans deux types de schémas de validation d'une étude source où certaines covariables sont mesurées avec des erreurs. Au lieu de se limiter à une famille restreinte de modèles jouissant de propriétés mathématiques commodes, nous suggérons que des considérations empiriques, adaptées aux données à traiter, guident les choix des modèles. La fonction de vraisemblance jointe de l'étude source, dans laquelle la mesure des covariables est sujette à erreur, et de l'étude de validation dans laquelle elle ne l'est pas, est maximisée et l'estimation et l'inférence sont dérivées en utilisant la théorie standard. Bien que la modélisation de l'erreur de mesure soit guidée par des considérations empiriques, la taille relativement réduite de l'étude de validation peut conduire à de mauvaises spécifications, avec comme conséquence des biais dans l'estimation et l'inférence sur la relation exposition-maladie. En utilisant un modèle non-paramétrique de l'erreur de mesure, les méthodes non-paramétriques suggérées par Robins et coll. (1994, 1995) sont exemptes des biais induits par des mauvaises spécifications du modèle d'erreur de mesure, au détriment naturellement de la puissance. Les deux approches paramétrique et non-paramétrique sont illustrées en détail sur l'exemple de la validation d'une étude des effets sur la santé d'une exposition professionnelle aux agents anti-cancéreux chez les pharmaciens (Valanis et coll., 1993). Un modèle à rapport de prévalence constant, avec des erreurs suivant une loi gamma, est établi et vérifié empiriquement sur les données disponibles. Une réanalyse des données, prenant en compte les erreurs de mesure, conduit à un triplement du risque relatif, sans diminution de la puissance. Les estimateurs semi-paramétriques sont cohérents avec les résultats de l'analyse paramétrique ce qui garantit que l'existence de biais importants dus à de mauvaises spécifications du modèle d'erreur de mesure est improbable.

References

Abramowitz, M. and Stegun, I. A. (eds). (1964). *Handbook of Mathematical Functions*. New York: Dover.

Armstrong, B. (1990). The effects of measurement errors in relative risk regressions. *American Journal of Epidemiology* **132,** 1176–1184.

Begun, J. M., Hall, W. J., Huang, W. M., and Wellner, J. A. (1983). Information and asymptotic efficiency in parametric–nonparametric models. *Annals of Statistics* **11,** 432–452.

Bischof, C., Carle, A., Corliss, G., Griewank, A., and Hovland, P. (1992). Adifor: Generating derivative codes from FORTRAN programs. *Scientific Programming* **1,** 11–29.

Carroll, R. J. and Stefanski, L. A. (1990). Approximate quasi-likelihood estimation in models with surrogate predictors. *Journal of the American Statistical Association* **85,** 652–663.

Carroll, R. J. and Wand, M. P. (1990). Semi-parametric estimation in logistic measurement error models. *Journal of the Royal Statistical Society, Series B* **53,** 573–587.

Carroll, R. J., Spiegelman, C. H., Lan, K. K. G., Bailey, K. T., and Abbott, R. D. (1984). On errors-in-variables for binary regression models. *Biometrika* **70,** 19–25.

Clayton, D. G. (1991). Models for the analysis of cohort and case-control studies with inaccurately measured exposures. In *Statistical Models for Longitudinal Data Analysis,* J. H. Dwyer, P. Lippert, M. Feinleib, and H. Hoffmeister (eds), 301–331. Oxford: Oxford University Press.

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society, Series B* **30,** 248–275.

Crouch, E. A. C. and Spiegelman, D. (1990). The evaluation of integrals of the form $\int f(t)\exp\{-t^2\}\,dt$. Application to logistic-normal models. *Journal of the American Statistical Association* **85,** 464–469.

Flanders, W. D. and Greenland, S. (1991). Analytic methods for two-stage case-control studies and other stratified designs. *Statistics in Medicine* **10,** 739–747.

Gong, G. and Samaniego, F. J. (1981). Pseudo maximum likelihood estimation: Theory and applications. *Annals of Statistics* **9,** 861–869.

Greenland, S. (1987). Interpretation and choice of effect measures in epidemiologic analyses. *American Journal of Epidemiology* **125,** 761–768.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite population. *Journal of the American Statistical Association* **47,** 663–685.

McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models,* 2nd edition. London: Chapman and Hall.

Moolgavkar, S. H. and Venzon, D. J. (1987). Confidence regions in curved exponential families: Application to matched case-control and survival studies with general relative risk functions. *Annals of Statistics* **15,** 346–359.

Panier, E. R. and Tits, A. L. (1993). On combining feasibility, descent and superlinear convergence in inequality constrained optimization. *Math Programming* **59,** 261–276.

Richardson, S. and Gilks, W. R. (1993). Conditional independence models for epidemiological studies with covariate measurement error. *Statistics in Medicine* **12,** 1703–1722.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89,** 846–866.

Robins, J. M., Hsieh, F., and Newey, W. (1995). Semi-parametric efficient estimation of a conditional density with missing or mismeasured covariates. *Journal of the Royal Statistical Society, Series B* **57,** 409–424.

Schafer, D. W. (1993). Likelihood analysis for probit regression with measurement errors. *Biometrika* **80,** 899–904.

Spiegelman, D. (1994). Cost-efficient study designs for relative risk modeling with covariate measurement error. *Journal of Statistical Planning and Inference* **42,** 187–208.

Thomas, D., Stram, D., and Dwyer, J. (1993). Exposure measurement error: Influence on exposure–disease relationships and methods of correction. *Annual Review of Public Health* **14,** 69–83.

Vaeth, M. (1985). On the use of Wald's test in exponential families. *International Statistical Review* **53,** 199–214.

Valanis, B. G., Vollmer, W. M., Labuhn, K. T., and Glass, A. G. (1993). Association of antineoplastic drug handling with acute adverse effects in pharmacy personnel. *American Journal of Pharmacists* **50,** 455–462.

Wacholder, S., Armstrong, B., and Hartge, P. (1993). Validation studies using an alloyed gold standard. *American Journal of Epidemiology* **137,** 1251–1258.

Whittemore, A. S. and Grosser, S. (1986). Regression methods for data with incomplete covariates. In *Modern Statistical Methods in Chronic Disease Epidemiology,* S. H. Moolgavkar and R. L. Prentice (eds), 19–35. New York: Wiley.

Whittemore, A. S. and Keller, J. B. (1988). Approximations for regression with covariate measurement error. *Journal of the American Statistical Association* **83,** 1057–1066.

Received August 1995; *revised March* 1996; *accepted August* 1996.

Appendix 1

Semi-Parametric Locally Efficient Estimation for a Binary Outcome with Random Validation Sampling

$\hat{\boldsymbol{\beta}}_{SPLE}$ is the solution to

$$U(\boldsymbol{\beta}) = U_1(\boldsymbol{\beta}) + U_2(\boldsymbol{\beta}) = 0, \tag{A1.1}$$

where

$$U_1(\boldsymbol{\beta}) = \sum_{i=1}^{n_2} \psi_i(\boldsymbol{\beta}),$$

where

$$\psi_i(\boldsymbol{\beta}) = \frac{\partial \log[f_1(D_i \mid x_i, U_{1i}; \boldsymbol{\beta})]}{\partial \boldsymbol{\beta}}$$

and

$$U_2(\boldsymbol{\beta}) = \sum_{i=1}^{n_1} \phi(D_i, X_i, U_i) - \frac{1-\pi}{\pi} \sum_{i=1}^{n_2} E_{D|x_i,U_{1i}}[\phi(D_i, X_i, U_i)],$$

where

$$\begin{aligned}
\phi(D_i, X_i, U_i) &= (I + \hat{m}_i)^{-1} \hat{Z}_i, \\
\hat{Z}_i &= [\hat{Z}_i(0), \hat{Z}_i(1)]', \\
\hat{Z}_i(d) &= \mathrm{E}_{x|d,X_i,U_i}[\psi(d \mid x, U_{1i}; \boldsymbol{\beta})], \\
\hat{m}_{rs,i} &= \frac{1-\pi}{\pi} E_{x|d=r,X_i,U_i}[f_1(D = s \mid x, U_{1i}; \boldsymbol{\beta})],
\end{aligned}$$

and

$$\mathrm{E}_{x|d,X,U}[h(x, U_1)] = \frac{\sum_{k=1}^{n_2} h(x, U_1) f_1(d \mid x_k, U_1; \boldsymbol{\beta}) f_2(x_k \mid X, U_2; \hat{\theta})}{\sum_{k=1}^{n_2} f_1(d \mid x_k, U_1; \boldsymbol{\beta}) f_2(x_k \mid X, U_2; \hat{\theta})},$$

where $\mathrm{E}_{x|d,X_i,U_i}$ denotes the expectation with respect to x conditional jointly on (d, X_i, U_i), and $f_2(x \mid X, U_2; \hat{\theta})$ is the parametric measurement error model fit to the data as described in Section 2.3.1.

Appendix 2

Derivation of $f_3(D = 1 \mid X, U)$ for the Pharmacists' Data

$$\begin{aligned}
f_3(D = 1 \mid X, U; \beta, \mu, \nu) &= \int_0^{+\infty} f_1(D = 1 \mid x, U_1, \beta) f_2(x \mid X, U_2, ; \mu, \nu)\, dx \\
&= \int_0^{+\infty} \left\{ e^{\beta_0 + \beta_x x + \beta'_{U_2} U_2} \frac{1}{\Gamma[\nu(X, U_{2\nu})]} x^{\nu(X,U_{2\nu})-1} \right. \\
&\qquad \left. \cdot \left[\frac{\nu(X, U_{2\nu})}{\mu(X, U_{2\mu})}\right]^{\nu(X,U_{2\nu})} e^{-\frac{\nu(X,U_{2\nu})x}{\mu(X,U_{2\mu})}} \right\} dx,
\end{aligned}$$

where $x > 0, \mu > 0, \nu > 0, U_2 = (U_{2\mu}, U_{2\nu})$, and $U =$ the unique elements of U_1 and U_2.

$$f_3(D=1 \mid X,U) = \left(\frac{\nu}{\mu}\right)^{\nu} \frac{1}{\Gamma(\nu)} e^{\beta_0+\beta'_{U_1}U_1} \int_0^{\infty} e^{-[(\nu/\mu)-\beta_x]x} x^{\nu-1}\, dx$$
$$= \left(\frac{\nu}{\mu}\right)^{\nu} \frac{1}{\Gamma(\nu)} e^{\beta_0+\beta'_{U_1}U_1} \int_0^{\infty} e^{-aX} X^n\, dx.$$

From the Table of Definite Integrals (Abramowitz and Stegun, 1964),

$$\int_0^{\infty} x^n e^{-aX}\, dx = \frac{\Gamma(n+1)}{a^{n+1}}.$$

Let $a = \nu/\mu - \beta_x$ and $n = \nu - 1$. Then,

$$f_3(D=1 \mid X,U) = \left(\frac{\nu}{\mu}\right)^{\nu} \frac{1}{\Gamma(\nu)} e^{\beta_0+\beta'_{U_1}U_1} \frac{\Gamma(\nu)}{\left(\frac{\nu}{\mu}-\beta_x\right)^{\nu}}$$
$$= e^{\beta_0+\beta'_{U_1}U_1} \left[\frac{\nu}{\mu}\left(\frac{\mu}{\nu-\mu\beta_x}\right)\right]^{\nu}$$
$$= e^{\beta_0+\beta'_{U_1}U_1} \left[\frac{\nu}{\nu-\mu\beta_x}\right]^{\nu},$$

where $\nu/\mu - \beta_x > 0$ and $\nu > 0$. Note that this derivation holds for arbitrary specifications of $\mu(X, U_{2\mu})$ and $\nu(X, U_{2\nu})$.

Appendix 3

Parametric Constraints on $L_{ACE}(\beta, \mu, \nu)$

$$\beta_0 + \beta_x x_i + \beta'_{U_1} U_{1i} \le 0 \quad \forall i, i = 1, \ldots, n_2, \tag{10a}$$
$$\mu_i = a + b'(X_i, U_{2\mu,i}) > 0 \quad \forall i, i = 1, \ldots, n_2, \tag{10b}$$
$$\nu_i = d + e'(X_i, U_{2\nu,i}) > 0 \quad \forall i, i = 1, \ldots, n_1 + n_2, \tag{10c}$$
$$\frac{\nu_i}{\mu_i} - \beta_x > 0 \quad \forall i, i = 1, \ldots, n_1, \tag{10d}$$
$$e^{\beta_0+\beta'_{U_1}U_{1i}} \left(\frac{\nu_i}{\nu_i - \beta_x \mu_i}\right)^{\nu_i} < 1 \quad \forall i, i = 1, \ldots, n_1. \tag{10e}$$