THE INSTITUTE OF STATISTICS

UNIVERSITY OF NORTH CAROLINA SYSTEM

META-ANALYSIS, MEASUREMENT ERROR AND CORRECTIONS FOR ATTENUATION

by

R. J. Carroll and L. A. Stefanski

Institute of Statistics Mimeograph Series No. 2201

October 1991

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

MIMEO # 2201 CARROLL, STEFANSKI
META ANALYSIS, MEASUREMENT ERROR
AND CORRECTIONS FOR ATTENUATION
1991

NAME	DATE

META–ANALYSIS, MEASUREMENT ERROR AND CORRECTIONS FOR ATTENUATION

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

L. A. Stefanski
Department of Statistics
North Carolina State University
Raleigh, NC 27695

ABSTRACT

MacMahon, et al. (1990) present a meta–analysis of the effect of blood pressure on coronary heart disease, as well as new methods for estimation in measurement error models for the case when a replicate or second measurement is made of the fallible predictor. The correction for attenuation used by these authors is compared to others already existing in the literature, as well as to a new instrumental variable method. The assumptions justifying the various methods are examined and their efficiencies are studied via simulation. Compared to the methods we discuss, that of MacMahon, et al. (1990) may have substantial bias in some circumstances because it does not take into account: (i) possible correlations among the predictors within a study; (ii) possible bias in the second measurement; or (iii) possibly differing marginal distributions of the predictors across studies.

1. INTRODUCTION

In this paper we discuss some problems that arise in regression meta-analysis when the predictor of interest is measured with nonnegligible error. Motivating our study is a recent paper by MacMahon, et al. (1990) presenting a meta-analysis of the effect of blood pressure on stroke and coronary heart disease. The methodology employed by MacMahon, et al. (1990) has been publicized in the press (Palca, 1990a,b), and it is likely to promote the use of measurement-error-model meta-analysis in future studies.

Underlying the methodology proposed by MacMahon, et al. (1990) is a correction for attenuation due to measurement error that employs both baseline and post-baseline measurements of blood pressure from one of the studies. We compare their correction for attenuation to some of the standard methods in the statistical literature for adjusting regression coefficients for the effects of measurement error. The assumptions justifying the various methods are examined and the estimated corrections for attenuation are studied via a small simulation experiment.

A necessary feature of a meta-analysis is that the individual studies share sufficient common ground to warrant the combination of information from the different data sets. With respect to measurement error modelling this imposes some restrictions on the relationship between the so-called true predictors and the surrogate predictors actually measured, especially when replicate measurements and/or validation data are limited. In this paper we work under the assumption that the measurement error variance is constant across studies. In the application of MacMahon, et al. (1990) this assumption seems reasonable, although there are likely to be other applications where the assumption is not warranted.

In $\oint 2$ we establish notation and discuss the effects of measurement error on regression meta-analyses. In $\oint 3$ we present and compare some corrections for attenuation due to measurement error. In $\oint 4$ we discuss the method employed by MacMahon, et al. (1990) in light of the results from $\oint 2$–3. We conclude in $\oint 5$.

2. GENERAL THEORY

2.1. Study Data, Models and Meta-Analysis in the Absence of Measurement Error

Let Y denote the response variate and X the risk factor that is to be investigated in the meta-analysis. We suppose the availability of data from $\mathcal{K}$ studies, with Y and X common to all studies. In addition, data from the k^{th} study includes a vector Z_k of study-specific covariates.

Justification for the meta-analysis depends on the assumption that the effect of X on Y is the same in all study populations after appropriate covariate adjustments. Thus as a model for the expected value of Y in the k^{th} population we have

$$E(Y \mid X, Z_k) = f(\alpha_k + \beta X + \gamma_k^T Z_k), \qquad k = 1, \ldots, \mathcal{K}. \tag{2.1}$$

The form of f is application dependent. For example, for linear regression f is the identity function, for binary regression f is commonly the logistic distribution function.

In some models, *e.g.*, logistic regression, (2.1) completely specifies the model once the type of variation, (Bernoulli, exponential, Poisson, *etc.*) in Y is specified. In other models there are additional variance parameters to model the residual variation which may be homoscedastic in the simplest cases or involve the mean function in addition to other variance parameters. Thus we allow for the possibility that accompanying (2.1) there is a variance-function model of the form

$$V(Y \mid X, Z_k) = \sigma_k^2 v(\alpha_k + \beta X + \gamma_k^T Z_k, \tau_k), \qquad k = 1, \ldots, \mathcal{K}. \tag{2.2}$$

Together (2.1) and (2.2) comprise a generalized quasilikelihood/variance-function model with linear effects, and includes, for example, generalized linear models and heteroscedastic nonlinear models; see Carroll & Ruppert (1988) for details.

Let $\widehat{\beta}_k$, $k = 1, \ldots, \mathcal{K}$ denote the estimates of β obtained by fitting the model to the k^{th} study data. The combined-study estimate of β is obtained as a weighted average

$$\widehat{\beta} = \sum_{k=1}^{\mathcal{K}} \widehat{w}_k \widehat{\beta}_k, \qquad (\widehat{w}_1 + \cdots + \widehat{w}_k = 1).$$

In the absence of special considerations, the natural choice of weights has $\widehat{w}_k \propto s_k^{-2}$ where s_k is the standard error of $\widehat{\beta}_k$.

2.2. The Effect of Measurement Error

Now suppose that the risk factor X is not accurately measured. Let W denote the measured risk factor and assume the representation $W = X + U$ where U is a measurement error independent of Y,

X and the other covariates. Furthermore assume that in the k^{th} study $\text{Var}(U) = \sigma_U^2$, independent of k.

This simple measurement error model is adequate for the epidemiologic application motivating the present paper. However, in some applications it may be necessary to allow for more complex dependencies between the proxy, W, and X and possibly the other covariates as well; see Fuller (1987) and Carroll & Stefanski (1991).

Now suppose that the models (2.1) and (2.2) are fit to the data with W substituted for X. The study-specific and combined-study estimators estimators so obtained are denoted $\tilde{\beta}_k$, $k = 1, \ldots, \mathcal{K}$ and $\tilde{\beta}$ respectively. In general neither $\tilde{\beta}_k$, $k = 1, \ldots, \mathcal{K}$ nor $\tilde{\beta}$ are consistent for β; see Stefanski (1985).

In the case that (2.1) and (2.2) define common homoscedastic linear models it is well known that $\tilde{\beta}_k$ consistently estimates $\lambda_k^{-1}\beta$ where

$$\lambda_k = \frac{\sigma_{W,k}^2}{\sigma_{W,k}^2 - \sigma_U^2}, \tag{2.3}$$

and $\sigma_{W,K}^2$ is the limiting value, assumed to exist, of $s_{W,k}^2 =$ the mean square error from the linear regression of W on Z_k in the k^{th} study data. Since $\sigma_{W,k}^2 = \sigma_{X,k}^2 + \sigma_U^2$ where $\sigma_{X,k}^2$ is the residual variation from the linear regression of X on Z_k, the attenuation factor for the k^{th} study group is

$$\lambda_k^{-1} = \frac{\sigma_{X,k}^2}{\sigma_{X,k}^2 + \sigma_U^2}. \tag{2.4}$$

It is evident from (2.4) that attenuation generally varies across study groups, and that colinearity between X and Z_k accentuates attenuation due to measurement error. Note that $\sigma_{W,k}^2 = \text{Var}(W)$ and $\sigma_{X,k}^2 = \text{Var}(X)$ only in the absence of covariates or when X and Z_k are uncorrelated.

It follows that for linear models the combined-study estimator $\tilde{\beta}$ converges in probability to $\lambda^{-1}\beta$ where

$$\lambda^{-1} = \sum_{k=1}^{\mathcal{K}} w_k \lambda_k^{-1},$$

where w_k are the limiting values of weights $\tilde{w}_k$.

Assuming that data are available for consistent estimation of λ_k, by say $\tilde{\lambda}_k$, $k = 1, \ldots, \mathcal{K}$, study-specific estimators of β corrected for attenuation are obtained as $\tilde{\beta}_{k,\text{CA}} = \tilde{\lambda}_k \tilde{\beta}$, $k = 1, \ldots, \mathcal{K}$. A combined-study estimator can be obtained by either taking a weighted average of $\{\tilde{\beta}_{k,\mathrm{CA}}\}$ or as $\tilde{\lambda}\tilde{\beta}$ where $\tilde{\lambda}$ is a consistent estimator of λ. The latter estimator is just a particular weighted average of $\{\tilde{\beta}_{k,\mathrm{CA}}\}$.

Starting with (2.1) and (2.2) and the simple measurement error model $W = X + U$ the following approximations to $E(Y|W)$ and $\mathrm{Var}(Y|W)$ can be derived:

$$E(Y|W, Z_k) \approx f\{\alpha_k + \beta E(X \mid W, Z_k) + \gamma_k^T Z_k\} \approx f(\alpha_k^* + \lambda_k^{-1}\beta W + \gamma_k^{*T} Z_k); \tag{2.5}$$

$$V(Y|W, Z_k) \approx \sigma_k^2 v\{\alpha_k + \beta E(X \mid W, Z_k) + \gamma_k^T Z_k, \tau_k\} \approx \sigma_k^2 v(\alpha_k^* + \lambda_k^{-1}\beta W + \gamma_k^{*T} Z_k, \tau_k). \tag{2.6}$$

The first approximation in (2.5) follows via a Taylor series expansion of (2.1) in X around $E(X \mid W, Z_k)$. The second approximation is obtained by replacing $E(X \mid W, Z_k)$ with the best linear approximation to this regression function. The approximations in (2.6) are similarly obtained. Estimation based on substituting $E(X \mid W, Z_k)$ for X is discussed by Carroll & Stefanski (1991), Gleser (1990), Fuller (1987), Pierce, et al. (1991), Prentice (1982), Rosner, et al. (1989, 1990) and Rudemo, et al. (1989), among others.

Equations (2.5) and (2.6) show that if the models (2.1) and (2.2) are fit to the data with W substituted for X, then the asymptotic bias in $\tilde{\beta}_k$ will be approximately the same as if linear models had been fit to the data. That is $\tilde{\beta}_k$ converges in probability to some value, say $\beta_k^\dagger$, that is approximately equal to $\lambda_k^{-1}\beta$. It follows that whenever the approximations in (2.5) and (2.6) are justified, the k^{th} study estimator can be corrected for attenuation just as in the case of linear models, *viz.*, $\tilde{\beta}_{k,\mathrm{CA}} = \tilde{\lambda}_k\tilde{\beta}_k$. Similarly a combined-study estimator corrected for attenuation is obtained as a weighted average of $\tilde{\beta}_{k,\mathrm{CA}}$, $k = 1, \ldots, \mathcal{K}$.

In certain applications, *e.g.*, logistic regression with $\Pr(Y = 1)$ near 1 or 0, the variation in $\tilde{\lambda}$ will be small relative to that in $\tilde{\beta}_k$. In such cases $\mathrm{Var}(\tilde{\beta}) \approx \sum_{k=1}^{\mathcal{K}} \tilde{w}_k^2 \tilde{\lambda}_k^2 \tilde{s}_k^2$ where $\tilde{s}_k$ is the usual standard error of $\tilde{\beta}_k$. In this case the best choice of weights has $\tilde{w}_k \propto (\tilde{\lambda}_k \tilde{s}_k)^{-2}$ for which the standard error of $\tilde{\beta}$ is approximately $\{\sum_{k=1}^{\mathcal{K}} (\tilde{\lambda}_k \tilde{s}_k)^{-2}\}^{-1/2}$.

3. ESTIMATING CORRECTIONS FOR ATTENUATION

3.1. Introduction

The composition of the data, especially with regard to replicate measurements, validation data and the presence of instrumental variables, generally dictates the appropriate method of estimating λ_k, $k = 1, \ldots, \mathcal{K}$. We will discuss only methods that are applicable to the epidemiologic application motivating the paper, MacMahon *et al.* (1990).

In that application all studies measured diastolic blood pressure (DBP) at baseline; this is W in the established notation. In addition, one study, Framingham, measured DBP at two-years and four-years post baseline. As in MacMahon, et al. (1990) we will use only the four-year post-baseline measurement. So now in addition to having available (Y, W, Z_k) in all k studies, there is available in one study, taken as the first, another variate, call it T. In the application under consideration T is the post-baseline measurement of DBP; more generally we regard T as a *second measurement* of X.

We make a distinction between a *second measurement* and a *replicate measurement.* The former implies only that T and X are correlated. The latter embodies the usual statistical notion of replicates; W and T are replicate measurements of X when $W = X + U_1$, $T = X + U_2$ and U_1 and U_2 are independent and identically distributed. The distinction is useful for it dictates when T should be employed as an instrumental variable; see Fuller (1987, p. 52).

The data from the first study contain most of the information for estimating σ_U^2. These data are used either to provide a direct estimate of σ_U^2 or an indirect estimate by first estimating λ_1 directly and then obtaining an estimate of σ_U^2 via (2.3), *viz.*,

$$\widehat{\sigma}_U^2 = \frac{\widehat{\lambda}_1 - 1}{\widehat{\lambda}_1} s_{W,k}^2. \tag{3.1}$$

In either case λ_k, $k = 2, \ldots, \mathcal{K}$ are estimated by

$$\tilde{\lambda}_k = \frac{s_{W,k}^2}{s_{W,k}^2 - \widehat{\sigma}_U^2}.$$

3.2. When W and T are Replicates

For the model with $W = X + U_1$ and $T = X + U_2$ where U_1 and U_2 are independent, identically distributed and independent of Z_1, the coefficient of W in the linear regression of T on (W, Z_1), denoted $\beta_{T|\underline{W},Z_1}$, is equal to λ_1^{-1}. Thus an estimator of λ_1 is

$$\overline{\lambda}_1 = \frac{1}{\widehat{\beta}_{T|\underline{W},Z_1}}. \tag{3.2}$$

Although statistical justification for this estimator is not strong, it has much intuitive appeal. In particular, it makes clear the connection between attenuation due to measurement error and the more widely understood phenomenon of regression to the mean. Apart from differences due to grouping and the inclusion of covariates, $\overline{\lambda}_1$ corresponds to the estimator employed by MacMahon, et al. (1990), see § 4 for additional discussion.

Objections to this estimator arise because it is not symmetric in T and W, while the statistical model is. Under the replication model is it natural to estimate σ_U^2 by

$$\widehat{\sigma}_U^2 = (1/2) \text{ sample variance of the } (W_i - T_i), \tag{3.3}$$

$\sigma_{W,k}^2$ by $s_{W,k}^2 =$ the mean square error from the linear regression of W_k on Z_k, and then λ_k by

$$\tilde{\lambda}_k = \frac{s_{W,k}^2}{s_{W,k}^2 - \widehat{\sigma}_U^2}, \qquad k = 1, \ldots, \mathcal{K}. \tag{3.4}$$

Even this procedure is not completely satisfying, for in the replication model the best measurement of X in the first study is $W^* = (T + W)/2$ and logic dictates first regressing Y_1 on W^* and Z_1 obtaining $\widehat{\beta}_1^*$ with corresponding attenuation factor

$$\lambda_1^{*-1} = \frac{\sigma_{X,1}^2}{\sigma_{X,1}^2 + \sigma_U^2/2}.$$

The measurement error variance, σ_U^2, again is estimated as in (3.3), and the mean square error from the regression of W^* on Z_1, denote it $s_{W,1}^{*2}$, consistently estimates $\sigma_{X,1}^2 + \sigma_U^2/2$. Thus

$$\tilde{\lambda}_1^* = \frac{s_{W,1}^{*2}}{s_{W,1}^{*2} - \widehat{\sigma}_U^2/2} \tag{3.5}$$

is a consistent estimator of λ_1^*.

The latter procedure makes more efficient use of the data but may be objected to on the grounds that it treats the first study differently from the rest. Greater similarity among studies, in this case having the study-specific analyses depend on baseline data only, makes it easier to present and defend the meta-analysis via nontechnical arguments.

Treating baseline and post-baseline measurements equally may be objectionable on statistical grounds as well. It would not be surprising to find differences in cohort distributions of DBP measurements taken four years apart, especially among study participants. In general, when the assumption that W and T are replicates is untenable, T should be employed as an instrumental variable.

3.3. T as an Instrumental Variable

Approximate instrumental variable estimation in generalized quasilikelihood/variance-function models will be studied in detail elsewhere. Following is a brief outline of that theory and a discussion of its application in meta-analysis of measurement error models.

If (2.1) and (2.2) are expanded around $E(X \mid T, Z_k)$ instead of $E(X \mid W, Z_k)$, then analogous to (2.5) and (2.6) we obtain, after replacing $E(X \mid T, Z_k)$ with its best linear approximant, the approximations

$$E(Y|T, Z_k) \approx f\{\alpha_k + \beta E(X \mid T, Z_k) + \gamma_k^T Z_k\} \approx f(\alpha_k^* + \delta_1 \beta T + \gamma_k^{*T} Z_k);$$

$$V(Y|T, Z_k) \approx \sigma_k^2 v\{\alpha_k + \beta E(X \mid T, Z_k) + \gamma_k^T Z_k, \tau_k\} \approx \sigma_k^2 v(\alpha_k^* + \delta_1 \beta T + \gamma_k^{*T} Z_k, \tau_k),$$

where δ_1 is the coefficient of T in the best linear approximation to the regression of X on T and Z_1.

Thus if the model (2.1) and (2.2) is fit to the data with T replacing X, the estimated coefficient of T, denoted $\widehat{\beta}_{Y|\underline{T},Z_1}$, is approximately consistent for $\delta_1\beta$.

Now for the additive model $W = X + U_1$, it is easy to establish that δ_1 is also the coefficient of T in the best linear approximation to the regression of W on T and Z_1, and thus can be consistently estimated by $\widehat{\beta}_{W|\underline{T},Z_1}$ = the estimated coefficient of T in the least squares regression of W on T and Z_1.

This leads to the approximate instrumental variable estimator

$$\tilde{\beta}_{1,IV} = \frac{\widehat{\beta}_{Y|\underline{T},Z_1}}{\widehat{\beta}_{W|\underline{T},Z_1}},$$

from which is derived the estimator of λ_1

$$\tilde{\lambda}_{1,IV} = \frac{\tilde{\beta}_{1,\mathrm{IV}}}{\tilde{\beta}_1}. \tag{3.6}$$

3.4. Comparing the Estimators of Attenuation

The four estimators $\overline{\lambda}_1$, $\tilde{\lambda}_1$, $\tilde{\lambda}_1^*$ and $\tilde{\lambda}_{1,IV}$, are all consistent for λ_1 when W and T are replicate measurements of X. We now examine the effect of departures from the replicate measurement error model on the four estimators.

We assume that $W = X + U_1$ as before, but that

$$T = \xi + \eta X + U_2^* \tag{3.7}$$

where U_2^* is a random error independent of U_1 and Z_1, but not necessarily having the same distribution as U_1. The model for T can be motivated as follows.

Let X and X_* denote the 'true' DBP at baseline and four years post-baseline respectively, of a randomly selected patient. If there is no change in true DBPs over the study period then $X = X_*$. If there is change, then it is reasonable to assume that X and X_* are jointly normal, in which case the regression of X_* on X is linear. Provided T is an unbiased measurement of the true DBP at follow-up, then $T = E(X_* \mid X) + U_2^*$ where $E(X_* \mid X) = \xi + \eta X$ and U_2^* is the sum of the measurement error and the residual error $X_* - E(X_* \mid X)$.

For this model $\eta = \rho\sigma_{X_*}\sigma_X^{-1}$, where $\rho = \text{corr}(X, X_*)$. Note that $\eta < 1$ unless $\text{Var}(X_*) \geq \rho^{-2}\text{Var}(X)$. So unless the variation in true DBPs increases over time, we expect $\eta < 1$.

Now consider $\overline{\lambda}_1$ defined in (3.2). For the model described above, $\beta_{T|\underline{W},Z_1}$ = the coefficient of W in the linear regression of T on (W, Z_1), is $\eta\lambda_1^{-1}$, and thus $\overline{\lambda}_1$ is a consistent estimator of λ_1/η. The correction for attenuation is overestimated in the common situation that $\eta < 1$ and underestimated when $\eta > 1$.

The second correction for attenuation, $\tilde{\lambda}_1$ given in (3.4) depends on the post-baseline measurements only via (3.3). If the T_i in (3.3) follow the model in (3.7) then $\widehat{\sigma}_U^2$ is a consistent estimator of

$$\frac{1}{2}\left\{\sigma_{U_1}^2 + \sigma_{U_2^*}^2 + (\eta - 1)^2\sigma_X^2\right\}.$$

This is greater than $\sigma_{U_1}^2$, and thus results in over correction for attenuation, when

$$\sigma_{U_2^*}^2 + (\eta - 1)^2\sigma_X^2 > \sigma_{U_1}^2. \tag{3.8}$$

For the model described above, U_1 is the measurement error at baseline and U_2^* is the sum of the measurement error at post-baseline and the residual error $X_* - E(X_* \mid X)$. Thus under constant measurement error variance, $\sigma^2_{U_2^*} > \sigma^2_{U_1}$ and the inequality in (3.8) holds.

The third correction for attenuation, $\tilde{\lambda}_1^*$ in (3.5), depends on the post-baseline measurements via $\hat{\sigma}^2_U$ as well as through $s^{*2}_{W,1}$, see (3.5). However, it propagates through to the the other studies only through $\hat{\sigma}^2_U$, which is generally overestimated under (3.7) resulting in positive bias in the corrections for attenuation in the other data sets, $k = 2, \ldots, \mathcal{K}$.

The instrumental variable estimator, $\tilde{\lambda}_{1,IV}$, depends only on a nonzero correlation between T and X and approximate linearity of certain regression functions and is therefore robust to departures from the replicate-measurements model in the direction of (3.7).

Summarizing these results we have: (i) all of the corrections for attenuation are consistent when the replicate-measurements model holds; (ii) only the instrumental-variable correction for attenuation is consistent under the more general second-measurement error model (3.7); and (iii) excluding the instrumental variable correction, the general effect of departures from the replicate-measurements model is to inflate the corrections for attenuation.

The consistency robustness of the instrumental-variable method to departures from the replicate-measurements model is obtained at the expense of greater finite-sample variability in the estimated corrections for attenuation. Table 1 displays the results of a simulation study designed to compare the four estimators $\overline{\lambda}_1$, $\tilde{\lambda}_1$, $\tilde{\lambda}_1^*$ and $\tilde{\lambda}_{1,IV}$ and the corresponding estimators of β. The four methods are designated RM (regression to the mean), MM and MM* (method of moments) and IV (instrumental variable) respectively.

The model for the simulation study was:

$$Y \mid X, Z \sim N(\alpha + \beta X + \gamma^T Z, 1), \quad \alpha = 0,\ \beta = 1,\ \gamma = (0,0)^T,$$

$$(X, Z)^T \sim N(0_{3\times 1}, I_3), \qquad W \mid X \sim N(X, \sigma^2_U),$$

$$T \mid X \sim N(\xi + \eta X, 1 - \rho^2 + \sigma^2_U), \quad \xi = 0,\ \eta = \rho.$$

The measurement error variance and ρ were investigated at two levels $\sigma^2_U = 0.25$, 1.00 and $\rho^2 = 1.00$, 0.90. Sample size was set at $n = 100$, and 500 independent data sets were generated at each factor level combination. Note that for methods RM, MM and IV, $\lambda = 1.25,\ 2.00$ when $\sigma_U^2 = 0.25,\ 1.00$ respectively, while for method MM*, $\lambda = 1.125,\ 1.50$ when $\sigma_U^2 = 0.25,\ 1.00$.

Monte Carlo means and mean squared errors are reported for the estimates of β and λ. The instrumental variable estimators of λ and β have generally smaller biases than the other estimators in the study. Furthermore the biases for the cases of $\rho^2 = 1.00$ and 0.90 are comparable, confirming the consistency robustness noted above. In terms of mean squared error $\tilde{\beta}_{IV}$ performs well also. However, the mean squared error of $\tilde{\lambda}_{IV}$ is consistently larger than the mean squared errors of the other three estimators of λ indicating greater sampling variability in this estimator.

4. DISCUSSION OF THE META-ANALYSIS OF CHD AND DBP

4.1. The Corrections in the Framingham Data

MacMahon, et al. (1990) and Kannel, et al. (1986) describe analysis of the Framingham study. We give here analyses which are meant to illustrate the issues discussed in the previous sections. The Framingham data consist of measurements at different examinations spaced two years apart, so that, for example, Exam 5 takes place 4 years after Exam 3. In what follows, we define X to be the true diastolic blood pressure (DBP).

We consider two possibilities for the definitions of W, T and Y:

- (*Follow–up Instrument*) W = DBP at Exam 3, T = DBP at Exam 5, Y = coronary heart disease (CHD) incidence in a 10 year follow–up from Exam 3;
- (*Precursor Instrument*) W_* = DBP at Exam 5, T_* = DBP at Exam 3, Y = CHD incidence in a 10 year follow–up from Exam 5.

The first situation (follow–up instrument) corresponds more closely to that described by MacMahon, et al. (1990), in the sense that they defined T as the four year follow–up measure of blood pressure. The other predictors Z used here are age (at Exam 3), serum cholesterol (at Exam 3) and cigarettes/day (at Exam 1). It might be preferable to use Exam 5 instead of Exam 3 for age and serum cholesterol in the second situation, but we have not done so in order to correspond as closely as possible to the analysis of MacMahon, et al. (1990), and because the results for DBP will likely not change significantly upon the redefinition, due to the lack of correlation between observed DBP and Z, see below.

In order to keep the discussion as simple as possible, we will ignore the possibility that the last two components of Z are measured with error, noting only that the regression method (3.2), the moments method (3.4) and the instrumental variables method (3.6) are all easily extended to handle more than one covariate measured with error.

The use of T_* as a precursor instrument (the second situation given above) most closely fits into the framework for instrumental variables discussed in the previous section. The use of T as a follow–up instrument (the first situation given above) in this problem presents some interesting issues. Due to censoring by death, T is an unusual "second" measurement, i.e., it is not possible to measure T in those subjects who die before Exam 5. We next discuss the effect that such censoring has on the methods discussed in the previous section. Let D be the indicator that a subject is alive 4 years after baseline ($D = 1$), calling such subjects uncensored. For the moments estimator, besides the assumptions we have discussed previously, suppose that (T, W) are approximately independent and identically distributed given $D = 1$, and W is unbiased for X. Then since CHD disease is "rare", we find that $\sigma_U^2 = \text{var}(W|X) \approx \text{var}(W|X, D = 1) \approx (1/2)\text{var}(T - W|X, D = 1) \approx (1/2)\text{var}(T - W|D = 1)$. Hence, under these circumstances, the moments method applied to the uncensored data yields an approximately consistent correction for attenuation. For the regression estimator, the previously discussed assumptions, along with the rare disease notion, indicates that $E(X|W, Z) = E(T|W, Z) \approx E(T|W, Z, D = 1)$, so that the regression method applied to the uncensored data yields an approximately consistent correction for attenuation. The issue is more subtle for the instrumental variable method, because the response Y is more directly involved in defining the correction. The regression of Y on (T, Z) is more difficult to interpret because it is a logistic regression on those who have remained uncensored for four years after baseline. In order for the instrumental variable method to make sense in this context, we need that if $\text{pr}(Y = 1|X, Z) \approx f\{\alpha_1 + \beta X + \gamma_1^T Z\}$, where $f(\cdot)$ is the logistic distribution function, then $\text{pr}(Y = 1|X, Z, D = 1) \approx f\{\alpha_1 + \beta X + \gamma_1^T Z\}$. This is reasonable if censoring within four years of baseline represents only a small fraction of the CHD cases over 10 years, which is problematic in this example; there were 174 cases in a 10 year follow–up from Exam 3 of those who were alive and free of CHD at Exam 3, but only 149 cases in a 10 year follow–up from Exam 5 among those who were alive and free of CHD at Exam 5.

Follow–up Instrument: For the moments and regression methods, we followed the same procedure as MacMahon, et al., namely deleting those who developed CHD or died in the four years post baseline. In the Framingham data, we found W had sample variance 140.80, while $s^2_{W,1} = 138.443$. This means, in effect, that the predictors Z are uncorrelated with W. The method of moments correction is $\tilde{\lambda}_1 = 1.69$ since the estimate of σ^2_U equals 56.39. If one replaces $s^2_{W,1}$ by the sample variance of W, then the moments correction becomes 1.67. In this particular case, with this particular set of covariates, ignoring the covariates has little effect.

The coefficient for W in regressing T on (Z, W) is .5872, leading to a regression correction of $\overline{\lambda}_1 = 1.70$, essentially the same as the moments correction.

The sample mean and variance of W are 82.16 and 140.80, respectively, while the sample mean and variance of T are 83.51 and 136.93. While the sample means are quite close in numerical value, with a sample size of $n = 1605$ subjects who were alive and disease free at Exams 3 and 5, and with $\widehat{\sigma}^2_U = 56.39$, the difference of the sample means is approximately 5 standard errors from zero under the replication model.

For the instrumental variable method, using only those who were alive and free of CHD at Exam 5, we find that $\widehat{\beta}_{W|T,Z} = .5958$. Among those alive and free of CHD at Exam 3, $\widehat{\beta}_{Y|W,Z} = .02439$. Among those alive at Exam 5, $\widehat{\beta}_{Y|T,Z} = .01988$. This yields a correction of $\tilde{\lambda}_{1,IV} = 1.37$.

Precursor Instrument: The moments correction does not change in this situation. The regression correction does change, but only in the third significant digit. For the instrumental variables method, $\widehat{\beta}_{Y|W_*,Z} = .01985$, $\widehat{\beta}_{Y|T_*,Z} = .01580$, and hence $\tilde{\lambda}_{1,IV} = 1.34$.

To summarize the results of this purely illustrative exercise, all methods yield similar corrections using either a precursor or a follow–up instrument. The moments and regression method corrections are essentially the same (≈ 1.70), while the instrumental variable method suggests a smaller change (≈ 1.35). Under an assumption of normality, the standard error of the moments correction estimate is approximately 0.043 under the replication model; the calculation uses standard variance and covariance formulae for sample variances. We do not have full access to the data and hence cannot verify the normality assumption. However, it does appear that the methods suggest corrections which are nontrivially different.

In order to understand whether the differences in the corrections, and hence in the parameter estimates, might be due to a systematic bias on one of the methods, we performed a small simulation study. We assumed that there were no additional variables Z, and we observe the response Y, the surrogate W and the second measurement T. The parameters in the simulation were chosen to correspond roughly to those reported in the Framingham data. The sample size was $n = 1605$, of whom in the Framingham data, 157 CHD cases developed within 10 years of Exam #3. It was noted above that $\text{var}(X) \approx 80$, $\text{var}(U) \approx 60$, $\lambda \approx 1.7$ and hence $\beta \approx .0244 \times 1.7$. We standardized so that X was normally distributed with mean 0.0 and variance 1.0. With this normalization, our simulations were constructed under the distributional assumptions that $W|X \sim N(X, \sigma_U^2)$ and $T|X \sim N(\xi + \eta X, 1 - \rho^2 + \sigma_U^2)$, with $\xi = 0$, $\eta = \rho$, $\sigma_U^2 = .75$, $\rho^2 = .9, 1.0$. The binary responses followed a logistic linear model with intercept -2.25 and slope 0.371. We also considered but do not report here results for slope parameters 0.1855, 0.742. The results of the simulation are given in Table 2, which is based on 250 repetitions of the experiment.

In reading Table 2, one should note that there is no "correct" value of λ, because in logistic regression the corrections are only approximate. However, Table 2 demonstrates that, in the set of circumstances defined by the simulation, the methods all are centered at about the same value, although the IV method is more variable. In terms of estimating the logistic slope β, the methods all perform approximately the same, both in terms of center and variability. Thus, the difference observed among the estimators from the Framingham data cannot be explained simply by systematic bias of any of the methods under a normal distribution for errors and predictors. It is, however, possible that the IV method is different from the others here because of its increased variability, although other explanations such as outliers cannot be ruled out.

Of course, one can overemphasize the estimation of λ, when β is the real parameter of interest. Translated to the standardized scale used in the simulations, we found that the moments and instrumental variable estimates differed by $\tilde{\beta}_1 - \tilde{\beta}_{1,IV} \approx (1.70 - 1.35) \times .0244 \times 80^{1/2} = 0.076$, while in our simulations, the standard deviation of the difference was 0.106, 0.101 for $\rho^2 = 0.90, 1.00$, respectively. This suggests that the various point estimates of β in the Framingham data are not statistically significantly different. In addition, these calculations point out the importance of attaching standard errors to parameter estimates; we have not done so here because of lack of access to the data. Formulas for asymptotic standard errors for the moments and regression estimates are found as special cases of the results of Carroll and Stefanski (1990).

4.2. The Method of MacMahon, et al. (1990) applied to Framingham Data

The method of MacMahon, et al. (1990) is closely related to the regression method (3.2) and is described as follows. Let C_1 and C_1 be extreme intervals, i.e., $C_1 = \{W \leq \text{lower bound}\}$ and $C_2 = \{W \geq \text{upper bound}\}$, and for $p = 1, 2$, define

$$\widehat{C}_{p1} = \frac{n_1^{-1}\sum_{i=1}^{n_1} W_i I(W_i \in C_p)}{n_1^{-1}\sum_{i=1}^{n_1} I(W_i \in C_p)};$$

$$\widehat{C}_{p2} = \frac{n_1^{-1}\sum_{i=1}^{n_1} T_i I(W_i \in C_p)}{n_1^{-1}\sum_{i=1}^{n_1} I(W_i \in C_p)}$$

Then their correction for attenuation estimate is

$$\widehat{\lambda}_{M,1} = \frac{\widehat{C}_{21} - \widehat{C}_{11}}{\widehat{C}_{22} - \widehat{C}_{12}}. \tag{4.1}$$

It is easily shown that when T and W are replicates as defined in ∮ 3.2, and X is independent of Z, then $\widehat{\lambda}_{M,1}$ consistently estimates λ_1, although we can show theoretically that it is generally inefficient. However, violation of these assumptions can cause inconsistent estimates.

For example, suppose that we continue to assume normality and continue to assume that T is a replicate, but now allow X and Z to be correlated. Then (4.1) estimates not λ_1 but instead estimates $\text{Var}(W)/\left\{\text{Var}(W) - \sigma_U^2\right\}$, leading to a correction which is generally too *small* since $\text{Var}(W) > \sigma_W^2$.

On the other hand, according to our analysis, the effect of T not being a replicate is that the correction (4.1) is generally too *large*, see ∮ 2.

In the Framingham data, using T as a follow–up instrument, W and Z are essentially uncorrelated, so that the correction (4.1) yields a reasonable estimate of a 60% correction in the ordinary logistic regression coefficients.

In summary, the correction (4.1) proposed by MacMahon, et al. is inefficient. Its appropriateness rests fundamentally on the two assumptions that T is a replicate and that X is unrelated to all the other covariates, assumptions which appear to be reasonable in the Framingham data. Significant violation of these assumptions means that this correction should not be used.

4.3. Meta–Analysis Using the MRFIT Study

For an illustration of meta–analysis, we use as a second study the MRFIT data, see Kannel, et al. (1986). MacMahon, et al. (1990) compute their correction (4.1) in the Framingham data, and then apply it to all the other data sets in their meta–analysis, including the MRFIT data. As illustrated by (2.3), this is not always appropriate since the correction for attenuation need not be constant across studies. Since we do not have access to the MRFIT data, we can only illustrate the danger of applying the correction for the first (Framingham) study to a second (MRFIT) study in a meta–analysis.

The following numbers are purely illustrative. We will ignore the covariates Z as the evidence suggests that they are at best weakly predictive of W, i.e., we will assume that X is independent of Z and that T is a replicate. Then, for Framingham men, the sample variance of W is 140.80, $\widehat{\sigma}_U^2 = 56.39$ from (3.3) and hence $\tilde{\lambda}_1 = 1.67$. For the MRFIT data, Kannel, et al. state that the sample variance of W is 110.25, so that $\tilde{\lambda}_2 = 2.04$. In other words, the fact that the predictor W is much less variable in the MRFIT study means that its regression coefficient should have been corrected by the factor 2.04, not 1.66 as suggested by MacMahon, et al., a 23% difference.

5. CONCLUSIONS

We have considered corrections for attenuation in semilinear regression models. There is a tendency for users of measurement error methods to base corrections for attenuation on the variance of the fallible covariate W. We have noted that the correction depends on the size of the measurement error σ_U^2 and on the size of the regression MSE, σ_W^2, from regressing the surrogate W on all the other predictors Z, see (2.3). Failing to take into account the presence of other covariates can lead to an undercorrection of the regression coefficients.

In correcting the results of a single study, we have distinguished between *replicate measurements* and *second measurements* of a fallible covariate. For replicates, the regression to the mean (RM) method (3.2), the method of moments (MM) methods (3.4)–(3.5) and the instrumental variable (IV) method (3.6) all yield consistent estimates of the attenuation.

For those cases that the second measurement is biased for the fallible covariate, only the IV method consistently estimates the attenuation. We have shown that in this case, the usual effect of using RM or MM is to *overestimate* the correction for attenuation.

We have also noted that the correction for attenuation proposed by MacMahon, et al. (1990) yields a consistent estimate only if the second measurement is a replicate *and* the true and surrogate predictors X and W are unrelated to all the other covariates Z. Failure of these assumptions generally causes under– and over–corrections, respectively.

We have also discussed the use of corrections for attenuation in meta–analysis. We have stressed in (2.3)–(2.4) that *the proper correction for attenuation may vary from study to study*, depending on the marginal distribution of the predictors through the regression MSE from regressing W on Z. Using the same correction for attenuation across studies generally leads to biased results. The direction of the bias is not predictable.

ACKNOWLEDGEMENT

Carroll's research was supported by a grant from the National Institutes of Health. We are indebted to Paul Sorlie for computing the Framingham numbers.

REFERENCES

Carroll, R. J. & Ruppert, D. (1988). *Transformation and Weighting in Regression.* Chapman & Hall, London.

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

Cornfield, J. (1962). Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure: A discriminant function analysis. *Federal Proceedings*, 21, 58–61.

Fuller, W. A. (1987). *Measurement Error Models.* Wiley, New York.

Gleser, L. J. (1990). Improvements of the naive approach to estimation in nonlinear errors-in-variables regression models. In *Statistical Analysis of Measurement Error Models and Application*, P. J. Brown and W. A. Fuller, editors. American Mathematics Society, Providence.

Kannel, W. B., Neaton, J. D., Wentworth, D., Thomas, H. E., Stamler, J., Hulley, S. B. & Kjelsberg, M. O. (1986). Overall and coronary heart disease mortality rates in relation to major risk factors in 325,348 men screened for MRFIT. *American Heart Journal*, 112, 825–836.

MacMahon, S., Peto, R., Cutler, J., Collins, R., Sorlie, P., Neaton, J., Abbott, R., Godwin, J., Dyer, A., & Stamler, J. (1990). Blood pressure, stroke and coronary heart disease: Part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias. *Lancet*, 335, 765–774.

Palca, J. (1990a). Getting to the heart of the cholesterol debate. *Science*, 24, 1170–71.

Palca, J. (1990b). Why statistics may understate the risk of heart disease. *Science*, 24, 1171.

Pierce, D. A., Stram, D. O., Vaeth, M., Schafer, D. (1990). Some insights into the errors in variables problem provided by consideration of radiation dose–response analyses for the A–bomo survivors. Preprint.

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. & Spiegelman, D. (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine*, 8, 1051-1070.

Rosner, B., Spiegelman, D. & Willett, W. C. (1990). Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *American Journal of Epidemiology*, 132, 734–745.

Rudemo, M., Ruppert, D. & Streibig, J. C. (1989). Random effect models in nonlinear regression with applications to bioassay. *Biometrics*, 45, 349-362.

Stefanski, L. A. (1985). The effect of measurement error on parameter estimation. *Biometrika*, 72, 583–592.

TABLE 1: COMPARISON OF ESTIMATORS

This simulation is described in § 3.4. Here "RM"denotes the regression to the mean method correction (3.2), "MM" and "MM*" denote the method of moments corrections (3.4) and (3.5), and "IV" is the instrumental variable correction (3.6). Table entries are Monte Carlo means (upper entry) and mean squared errors (lower entry).

		$\sigma_U^2 = 0.25$				$\sigma_U^2 = 1.00$			
		RM	MM	MM*	IV	RM	MM	MM*	IV
$\rho^2 = 1.00$	β	1.005	1.009	1.003	1.001	1.033	1.053	1.027	1.024
		.018	.018	.015	.017	.060	.074	.041	.050
	λ	1.253	1.255	1.127	1.250	2.062	2.089	1.529	2.071
		.010	.004	.001	.013	.216	.188	.037	.344
$\rho^2 = 0.90$	β	1.062	1.065	1.032	1.002	1.097	1.120	1.061	1.027
		.026	.026	.018	.019	.091	.111	.055	.057
	λ	1.323	1.325	1.161	1.250	2.190	2.223	1.592	2.079
		.021	.012	.003	.016	.347	.317	.062	.375

TABLE 2: COMPARISON OF ESTIMATORS IN LOGISTIC REGRESSION

This simulation is described in ∮4.1. Here "RM"denotes the regression to the mean method correction (3.2), "MM" denotes the method of moments correction (3.4), and "IV" is the instrumental variable correction (3.6). The term "MAD" refers to the median absolute deviation from the median, "MSE" is mean squared error, "Mean AE" is the mean absolute error and "Median AE" is the median absolute error. The correct value of $\beta = .371$.

Results for λ				
	$\rho^2 = 1.00$	MM	RM	IV
Mean		1.756	1.755	1.780
Median		1.754	1.756	1.670
S.D.		.063	.065	.543
MAD		.045	.046	.323
	$\rho^2 = 0.90$	MM	RM	IV
Mean		1.851	1.846	1.775
Median		1.846	1.839	1.668
S.D.		.070	.071	.678
MAD		.049	.045	.290
Results for $\beta = .371$				
	$\rho^2 = 1.00$	MM	RM	IV
Mean		.386	.386	.376
Median		.389	.389	.377
MSE		.014	.014	.013
Mean AE		.094	.093	.090
Median AE		.081	.079	.079
	$\rho^2 = 0.90$	MM	RM	IV
Mean		.409	.408	.374
Median		.406	.409	.369
MSE		.015	.015	.013
Mean AE		.099	.097	.090
Median AE		.082	.076	.075

FOR THE REFEREES

6. CALCULATION OF STANDARD ERROR FOR MOMENTS ESTIMATE

Let $Y_1 = 2^{-1/2}(W - T)$ and $Y_2 = W$. Let s_{11} and s_{22} be the sample covariance matrices for the Y_1 and Y_2 observations. Then

$$\sigma_{11} = \sigma_U^2;$$
$$\sigma_{22} = \sigma_W^2,$$

and

$$\sigma_{12} = \sigma_{21} = \mathrm{cov}(Y_1, Y_2) = 2^{-1/2}\sigma_U^2.$$

According to Fuller's Appendix to Chapter 1, this means that

$$\mathrm{cov}(s_{11}, s_{11}) = 2\sigma_{11}^2;$$
$$\mathrm{cov}(s_{22}, s_{22}) = 2\sigma_{22}^2;$$
$$\mathrm{cov}(s_{11}, s_{22}) = 2\sigma_{12}^2 = \sigma_U^4.$$

If

$$\widehat{\lambda} = \frac{\widehat{\sigma}_W^2}{\widehat{\sigma}_W^2 - \widehat{\sigma}_U^2},$$

then

$$\widehat{\lambda} - \lambda \approx (1 - \lambda)\frac{\widehat{\sigma}_W^2 - \sigma_W^2}{\sigma_X^2} + \lambda\frac{\widehat{\sigma}_U^2 - \sigma_U^2}{\sigma_X^2}.$$

Hence,

$$n\mathrm{var}\left(\widehat{\lambda} - \lambda\right) \approx 2\frac{(1-\lambda)^2\sigma_W^4 + \lambda^2\sigma_U^4 + \lambda(1-\lambda)\sigma_U^4}{\sigma_X^4}.$$

Because $\sigma_U^2/\sigma_X^2 = \lambda - 1$ and $\lambda = \sigma_W^2/\sigma_X^2$, we thus have

$$\mathrm{var}\left(\widehat{\lambda} - \lambda\right) \approx 2\lambda^2(\lambda - 1)^2\{2 + (1-\lambda)/\lambda\}/n.$$

With $\lambda = 1.6$ as in the Framingham study, and $n = 1605$, the standard error for $\widehat{\lambda}$ is approximately 0.043.