ENVIRONMETRICS, VOL. 7, 441–451 (1996)

CAUSALITY, MEASUREMENT ERROR AND MULTICOLLINEARITY IN EPIDEMIOLOGY

JAMES V. ZIDEK, HUBERT WONG

Department of Statistics, University of British Columbia, 6356 Agriculture Road, Vancouver, Canada V6T 1Z2

NHU D. LE

Division of Epidemiology, British Columbia Cancer Agency, 600 West 10th Ave, Vancouver, Canada V5Z 4E6

AND

RICK BURNETT

Environmental Health Center, Health Canada, Tunney's Pasture, Ottawa, Canada K1A 0L2

SUMMARY

This paper demonstrates that measurement error can conspire with multicollinearity among explanatory variables to mislead an investigator. A causal variable measured with error may be overlooked and its significance transferred to a surrogate. The latter's significance can then be entirely spurious, in that controlling it will not predictably change the response variable. In epidemiological research, such a response may be a health outcome. The phenomenon we study is demonstrated through simulation experiments involving non-linear regression models. Also, the paper presents the measurement error problem in a theoretical setting. The paper concludes by echoing the familiar warning against making conclusions about causality from a multiple regression analysis, in this case because of the phenomenon presented in the paper.

KEY WORDS: errors in variables; multicollinearity; causality; Poisson regression; longitudinal analysis; environment

1. INTRODUCTION

This paper explores in non-linear regression analysis, the interaction of measurement error and multicollinearity. Our simulation studies show how their combined effect can mislead the investigator. Causes may be obscured. The wrong explanatory variables may be implicated by statistical significance. To be more precise, suppose that in an observational study a causal variable x is measured with a large error to yield X. A second regressor variable w is measured with small error as W. The variables w and x are highly correlated (collinear). The outcome of interest is Y. We show in Section 2 that the variation in Y explained by X can be statistically insignificant so that the investigator fails to discover the true role of x. At the same time, s/he may well find W to be significant even though that finding would be totally spurious.

The subject of this paper has been explored for linear regression (Fuller 1987, p. 55). We will review results for that case in the next section.

Our results may be of value to epidemiologists who commonly use non-linear regression in observational studies to identify possible causes of adverse health outcomes. We know such

CCC 1180–4009/96/040441–11

Received July 1995
Revised April 1996

investigators cannot prove cause. They cannot rule out the existence of confounding factors when they find a statistically significant association between the outcome and an explanatory variable. We show in the following that significant association implies the predictability of the outcome from that explanatory variable. However, that variable cannot be regarded as a design variable; changing its level need not yield a corresponding change in the outcome. For example, reducing daily average ground level ozone concentrations may not reduce daily hospital admission counts for acute respiratory morbidity; the strong statistical association between ozone levels and such hospital admissions (see Burnett and Krewski 1994) are based on observational, not designed studies. None the less, such findings would lead us to predict an increase in hospital admissions following a natural increase in ozone levels.

Even a designed study can mislead investigators seeking causal predictions. For example, Simpson's paradox occurs in multi-centre randomized trials of a treatment against a control when sample sizes are too small or randomization protocols are too non-uniform (Zidek 1984). Investigators may then erroneously conclude from aggregated multi-centre data that the treatment will predictably yield better results than the control.

Yet when they analyse the same data centre by centre, the controls can prove to be superior to the treatment in every single case. In observational studies, analogous difficulties due to aggregation express themselves through the ecological correlation fallacy (Robinson 1950).

However, we will not be concerned with that fallacy. Instead we focus on confounding factors, the main impediments to proving cause in observational studies.

In his celebrated example, Charlier showed that over a period of 40 years annual stork population totals in Oslo were correlated with the number of human births in that city ($r = 0{\cdot}86$). However, slaughtering storks would not have been an effective method of birth control since the true explanation for the association lay in a confounding factor, the size of the human population.

In observational studies, 'association', causal or otherwise, may be interpreted in terms of the outcome's predictability. Consider in general, a study population of items with measurable study variables; Y represents the measured response, X, x, W, w and so on, the regressor variables. The latter represent potential causes. They can be related to one another as well as subject to measurement errors.

Now if w were not known, the constant function, $E[g(Y)]$ which ignores w would seem to be a natural predictor of $g(Y)$, g denoting a given function; here and in the following, $E[\cdot]$ denotes the study population average. However, if w were available we might be able to find a better predictor than the constant function. That predictor would have the form $h(w)$ for some h. By 'better' we mean $E\{[g(Y) - h(w)]^2\} < E(\{g(Y) - E[g(Y)]\}^2)$. If there does exist a function g having such a better predictor, we call Y and w 'related' and otherwise 'unrelated'.

We can readily show that being 'unrelated' then means, $P(w \in A, Y \in B) = P(w \in A)$ $P(Y \in B)$ or $P(Y \in B | w \in A) = P(Y \in B)$ for all events A and B. In probability theory, a pair of random variables, w and Y satisfying the last condition are called 'independent' and otherwise 'dependent'. Note that independence obtains above even if we substitute an arbitrary non-negative measure of accuracy, $\rho\{\cdot, \cdot\}$ for $\{\cdot - \cdot\}^2$ (provided that $\rho[u, v] = 0$ if and only if $u = v$). In non-linear regression, we use the strong concept of predictive association implied by stochastic dependence. Indeed in the simulation study of Section 2, Y represents a Poisson count and the lack of association with the explanatory variable X will mean $P(Y = y | X = x_0) = P(Y = y)$ for all x_0.

When Y represents a continuous outcome, a weaker notion of non-association can be used, that requiring the g's and h's in the above definition to be linear. We then find that w and Y must be 'uncorrelated', i.e. linearly non-related. Except when these variables have a joint normal

distribution, linear association and association are not equivalent.

Between association and linear association are 'quadratic association' and a family of other types of association. Inexplicably, these intermediate notions of association have not been systematically explored. We will not pursue them here.

2. REVIEW OF RELATED THEORY

Fuller (1987) gives a comprehensive account of the theory of measurement error for linear models. That for non-linear models is under current development. Carroll and Stefanski (1990) give an account of that development within the quasi-likelihood framework. Measurement error in non-linear regression with random effects has been treated in unpublished work by the authors and their co-investigators. They assume small measurement errors and use Taylor expansion methods to deal with random cluster effects in a generalized estimating equations framework for longitudinal analysis. However, none of the existing theory enables us to accurately assess the effect of the phenomenon central to this paper, the combined effect of measurement error and multicollinearity.

The classical measurement error model may be described in the notation of Fuller (1987) by

$$Y_t = \beta_0 + \beta_1 x_t + \epsilon_t, X_t = x_t + u_t, t = 1, \ldots, n. \tag{1}$$

The $\{x_t\}$ may be regarded as fixed or random. Often $\epsilon_t \sim \text{NID}(0, \sigma_{ee})$, $t = 1, \ldots, n$ and $u_t \sim \text{NID}(0, \sigma_{uu})$, $t = 1, \ldots, n$ are assumed. Also, the $\{u_t\}$ and $\{x_t\}$ are assumed to be uncorrelated. The data consist of the pairs of measurements made on items drawn from the study population: $\{(Y_t, X_t) : t = 1, \ldots, n\}$. Adopting a common practice, an investigor might fit a linear model relating Y_t to X_t by ignoring the measurement in the explanatory variable. In that case s/he would obtain as an estimator for $\beta_1, \hat{\beta}_1 = \{\Sigma(X_t - \bar{X})^2\}^{-1}\Sigma(X_t - \bar{X})Y_t$. We find that the estimator is biased when averaged over the distribution described by the sampling model (1) which incorporates the measurement errors:

$$E\hat{\beta}_1 = \kappa_{xx}\beta_1. \tag{2}$$

Here $\kappa_{xx} = \sigma_{xx}/\sigma_{XX}, \sigma_{XX} = \sigma_{xx} + \sigma_{uu}$ and in general σ_{st} represents the study population covariance between any two study variables s and t. Fuller (1987) calls κ_{xx} the 'reliability ratio' which attenuates the regression coefficient. Even though β_1 has large magnitude, its least squares estimator can approach 0 when measurement error is large.

Measurement error also leads to an attentuation of the squared study population correlation between x and Y, again by the reliability ratio acting linearly. Thus measurement error can lead to the failure of an investigation to discover the true relationship between the outcome represented by Y and the explanatory variable represented by x when measurement error is present.

The failure of least squares analysis in the presence of measurement error stems technically from the non-linearity of $\hat{\beta}_1$ in the $\{x_t\}$. One might ask whether measurement error can be similarly misleading in a designed, as opposed to obervational, study. This leads us to the model for error proposed by Berkson (1950):

$$Y_t = \beta_0 + \beta_1 x_t + \epsilon_t, x_t = X_t - u_t, t = 1, \ldots, n. \tag{3}$$

The experimenter fixes the $\{X_t\}$ but the $\{x_t\}$ represent the actual levels at which the experiments are run due to uncontrolled variation. The $\{u_t\}$ and $\{x_t\}$ satisfy the same conditions as above.

Formally (1) and (3) look identical. However, they are not the same. Whereas previously the

$\{x_t\}$ and $\{u_t\}$ were uncorrelated, now they are not, and it turns out for the simple models under consideration, the investigator will not be mislead. S/he can use the least squares estimator for β_1 obtained by fitting the model $Y_t = b_0 + b_1 X_t$. That estimator will be unbiased.

For more complicated models, for example, those in which dependence on the $\{X_t\}$ is non-linear, this unbiasedness of coefficient estimators no longer obtains. So in summary, investigators need to recognize the deleterious nature of measurement error both in design and analysis, where in the latter case it may be possible to adjust estimators to obtain approximate unbiasedness at least.

Let us now turn to the central issue of concern in this paper and introduce another regressor variable W for which $\sigma_{xW} \neq 0$ while measurement model (1) holds. Although the co-ordinates of W like those of x can be fixed, or a combination of fixed and random, we assume for expository simplicity that X, x and W are all random. The investigator cannot rule out *a priori*, the possibility that W has an important explanatory role. So s/he would fit the model $Y_t = b_0 + b_1 X_t + b_2 W_t$ by least squares. Usually the resulting b_1 will not be an unbiased estimator of β_1, but it might be hoped that the fitted coefficient for W would not be significantly different from zero. This may well fail to be true also. As Fuller (1987, p. 55) observes:

> 'If theory specifies Y to be a function of x only, x is measured imperfectly by X, and W is correlated with x, then the coefficient for W in the multiple regression of Y on X and W is not zero.'

To get more insight into the difficulties involved in the situation described in the last paragraph, consider the study population coefficients of X and W under the measurement model (1) with the added assumption that W is independent of both u and ϵ. So we regard $(Y_t, X_t, W_t), t = 1, \ldots, n$ as a set of trivariate measurements made on a random sample of items drawn from the study population. Then under multivariate normal theory, the conditional expectation of Y_t given X_t and W_t is, disregarding an additive constant, $\delta W_t + \gamma W_t$; here

$$\delta = \Delta^{-1}(\sigma_{YX} - \beta_1 \pi_{22}^2 \sigma_{WW})$$

$$\gamma = -\pi_{22}\Delta^{-1}(\sigma_{YX} - \beta_1 \sigma_{XX})$$

$\Delta = \sigma_{XX} - \pi_{22}^2 \sigma_{WW}$ and $\pi_{22} = \sigma_{xW}/\sigma_{WW}$. The least squares estimator will be an unbiased estimator of β_1 if and only if $\sigma_{YX} = \beta_1 \sigma_{XX}$. The last condition holds if and only if $\gamma = 0$. In particular, the condition for unbiasedness cannot obtain under the realistic conditions, $\sigma_{ex} = \sigma_{ux} = \sigma_{eu} = 0$, for then $\sigma_{YX} = \beta_1 \sigma_{xx}$ while $\sigma_{XX} = \sigma_{xx} + \sigma_{uu}$ so that $\sigma_{YX} = \beta_1 \sigma_{XX} = -\beta_1 \sigma_{uu} < 0$ for non-negligible measurement error. Consequently

$$\gamma = \rho \beta_1 (\sigma_{xx}/\sigma_{WW})^{1/2}[1 + (1 - \rho^2)(\sigma_{xx}/\sigma_{uu})]^{-1} \tag{4}$$

where ρ denotes the correlation between x and W. Thus, when $\rho = 0$, $\gamma = 0$, but as $\rho \to 1$, $\gamma \to \beta_1$ if W has the same scale as x, i.e. $\sigma_{xx} = \sigma_{WW}$. The result obtains without regard to the level of measurement error and shows that as $\rho \to 1$, W will come to fulfil the role of a surrogate for x.

Equation (4) reveals something rather unexpected; as measurement error diminishes and $\sigma_{uu} \to 0$, $\gamma \to 0$ whatever be the values of the other parameters in the measurement/sampling model. W's surrogacy role will then be completely lost.

Corresponding results obtain for the other coefficient δ defined above, but for brevity these will not be presented. Results for more general measurement error models where Y is a continuous

response or vector of continuous responses could be developed with the help of the theory presented by Fuller (1987). However, we need models like that for Poisson regression where his theory proves of no avail. Indeed, existing theory such as Carroll and Stefanski's cited above, tends to be heavily computational. In more complex cases, it seems impossible to obtain the qualitative insight we get above for the simplest models. Thus in the next section we turn to simulation experiments to gain some insight into the joint effects of multicollinearity and measurement error in non-linear regression.

3. SIMULATION STUDY

In this section we present simulation results assessing the interaction between errors in variables and multicollinearity in Poisson regression. Two latent predictor variables, x and w, are drawn from a bivariate normal distribution with zero means, unit variances and correlation ρ. These two variables are observed with error as $X = x + u$ and $W = w + v$, where u and v are independent (of x and w and of each other) and normally distributed with zero means and standard deviations σ_{uu} and σ_{vv}, respectively. The response variable Y is Poisson distributed with mean $\exp(\alpha_0 + \alpha_1 x)$ where $\alpha_0 = 0$ and $\alpha_1 = 1$. For a range of values of ρ, σ_{uu} and σ_{vv}, we fit four Poisson regressions models:

1. Y on x alone;
2. Y on the pair (x, w);
3. Y on X alone, and
4. Y on the pair (X, W).

3.1. Simulation algorithm

The simulations proceeded as follows:

1. Fix the values of σ_{vv}, σ_{uu} and ρ.
2. Generate 100 pairs (x_i, w_i).
3. Generate 100 pairs (u_i, v_i) and add these to (x_i, w_i) to obtain (X_i, W_i).
4. Generate Poisson counts Y_i with means x_i.
5. Perform Poisson regression of Y assuming
   (a) $E(Y) = \exp(\alpha_0^* + \alpha_1^* x)$
   (b) $E(Y) = \exp(\beta_0^* + \beta_1^* x + \beta_2^* w)$
   (c) $E(Y) = \exp(\alpha_0 + \alpha_1 X)$
   (d) $E(Y) = \exp(\beta_0 + \beta_1 X + \beta_2 W)$.
   Record the coefficients and the corresponding asymptotic standard errors (based on the expected Fisher information adjusted by the scale estimate) from each fit.
6. Repeat steps 2–5 400 times.
7. Change ρ and repeat steps 2–6.
8. Change σ_{uu} and repeat steps 2–7.
9. Change σ_{vv} and repeat steps 2–8.
10. From each set of 400 replicate fits, compute the mean, sample standard error, and mean asymptotic standard error for each coefficient.

The algorithm was implemented in S-plus with model fits obtained using the 'Glim' function.

3.2. Simulation results

Results from our regressions on the true predictor values agreed with theory and will not be

presented. The following results seem more instructive.

Figure 1 displays the means for $\hat{\alpha}_1$, $\hat{\beta}_1$ and $\hat{\beta}_2$ computed from each set of 400 replicate fits as a function of σ_{uu} for values of ρ of 0, 0·5 or 0·9 and $\sigma_{vv} = 0$ (no error in w) or $\sigma_{vv} = 1$ (large error in w). Figures 2 and 3 display the corresponding sample and asymptotic standard errors for $\hat{\beta}_1$ and $\hat{\beta}_2$.

Observe the well-known phenomenon of $\hat{\alpha}_1$'s attenuation due to error in x. The theoretical value of α_1 is 1. As the error in measuring x increases (and σ_{uu} increases away from 0 towards its maximum value of 1 in our study), the value of α_1 decreases. The analyst would see that value declining in significance and eventually (beyond the range of our plots) the error bands (plotted in Figure 1) would encompass 0. At that point, the analyst would conclude that x did not significantly explain Y's variation even though we know that in this simulation study at least,

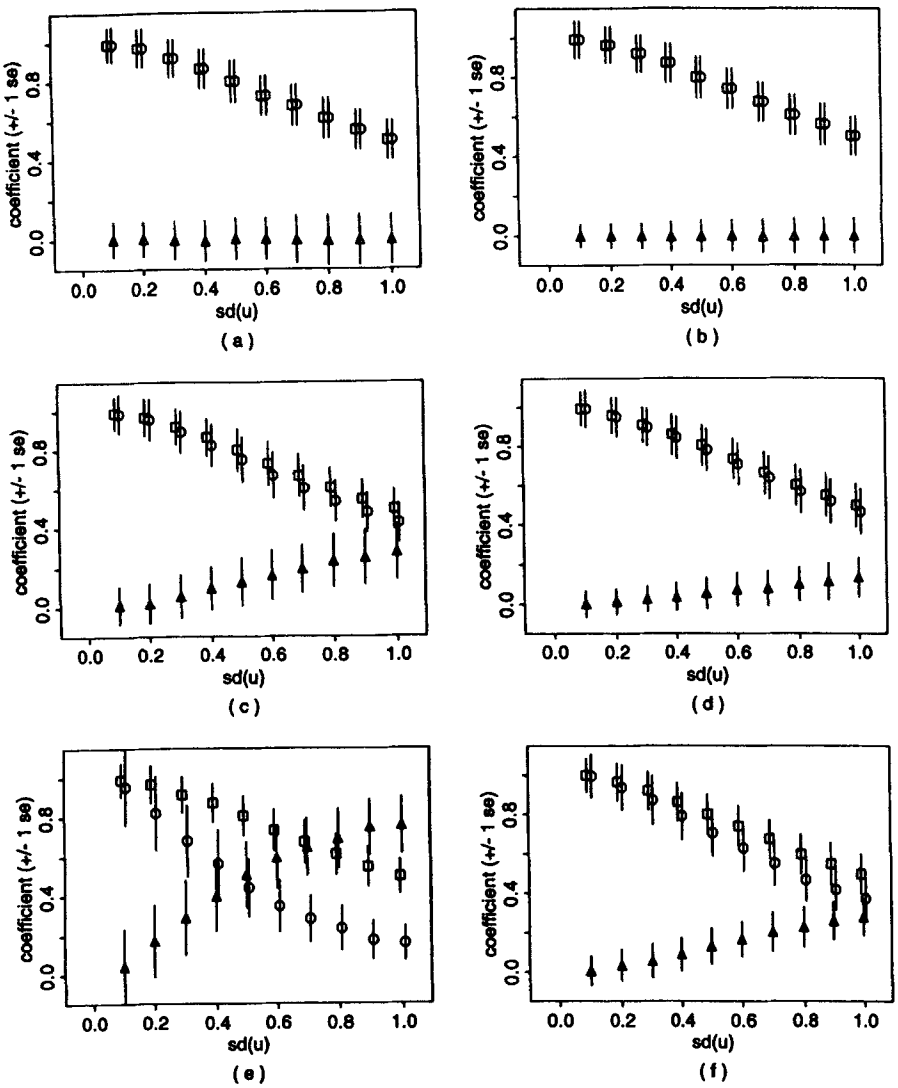

Figure 1. Averages with error bands for 400 non-linear regression coefficient estimates (α_1 □, β_1 ○, β_2 ▲) from samples of size 100 obtained under various levels of measurement error and multicollinearity: (*a*) $\rho = 0$, $\sigma(v) = 0$; (*b*) $\rho = 0$, $\sigma(v) = 1$; (*c*) $\rho = 0{\cdot}5$, $\sigma(v) = 0$; (*d*) $\rho = 0{\cdot}5$, $\sigma(v) = 1$; (*e*) $\rho = 0{\cdot}9$, $\sigma(v) = 0$; (*f*) $\rho = 0{\cdot}9$, $\sigma(v) = 1$

variation in x actually causes the broad scale variation in Y.

Our next observation concerns the anticipated transfer of effects from $\hat{\beta}_1$ to $\hat{\beta}_2$ due to ρ. Notice in Figure 1 that as ρ increases from 0 to 0·9, the value of $\hat{\beta}_2$ increases for any given level of measurement error in x. By the time the correlation between x and w, ρ, reaches 0·9, the paths traced out by the $\hat{\beta}$'s with changing σ_{uu} have crossed. Now β_2 has come to dominate β_1 even though the association between w and Y, like that between storks and babies mentioned in the introduction, is completely spurious.

For this transfer of apparent causality to occur, we need a high dose of both measurement error in x and association between x and w. Again looking at Figure 1, we see that at a value of $\rho = 0{\cdot}5$ the paths referred to in the last paragraph only contact each other when the error in

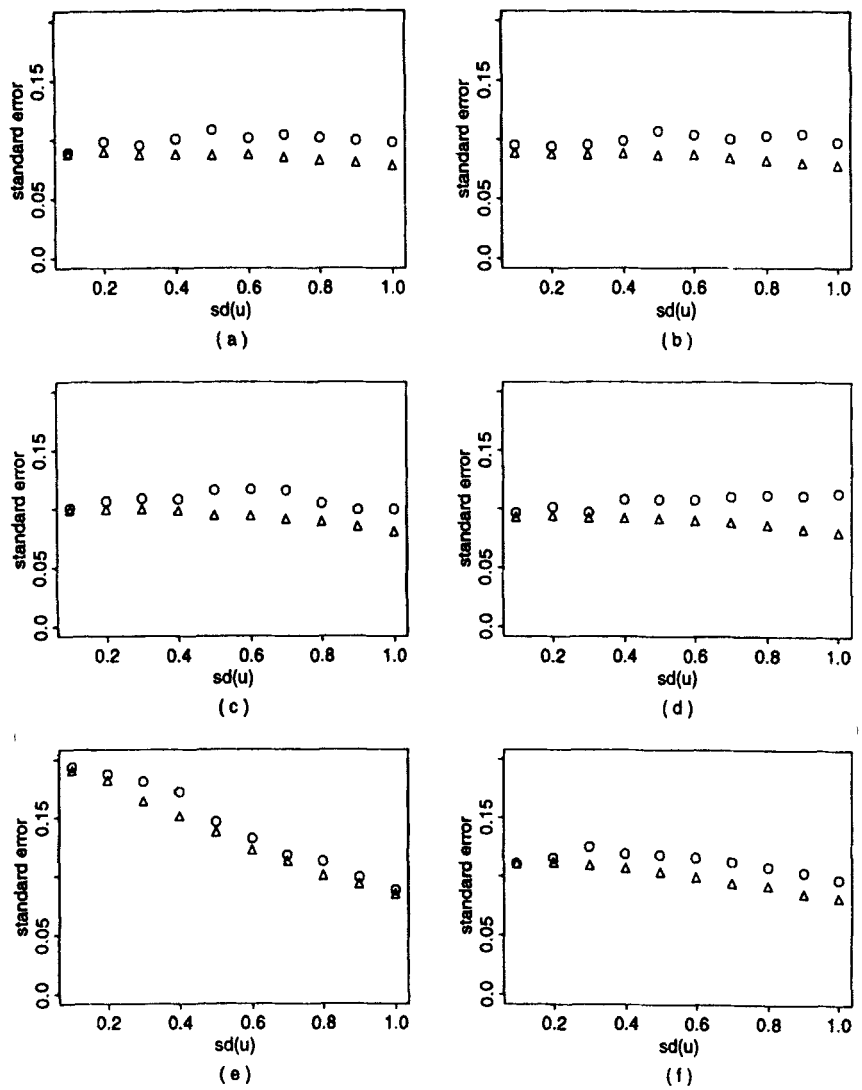

Figure 2. Standard errors (sample O, average asymptotic Δ) for β_1 obtained from 400 non-linear regressions with samples of size 100 under various levels of measurement error and multicollinearity: (*a*) $\rho = 0$, $\sigma(v) = 0$; (*b*) $\rho = 0$, $\sigma(v) = 1$; (*c*) $\rho = 0{\cdot}5$, $\sigma(v) = 0$; (*d*) $\rho = 0{\cdot}5$, $\sigma(v) = 1$; (*e*) $\rho = 0{\cdot}9$, $\sigma(v) = 0$; (*f*) $\rho = 0{\cdot}9$, $\sigma(v) = 1$

measuring x approaches 100 per cent of the simulated variation of x itself.

However, the analyst would come to see w as significant long before that extreme level of measurement error is seen. The error band for β_2 ceases to include 0 by the time the level of σ_{uu} has reached about 0·4; again the analyst would be deceived by the effect of measurement error into making an incorrect conclusion.

Figure 1 shows that the transfer-of-causality effect noted above is mitigated when w also becomes subject to error. The growth in the significance of β_2's estimate as a function of the error in x's measurement is retarded. However, it perversely 'borrows from strength' and exploits its strong association with the true cause of Y's variation, x, to become strongly significant even when $\sigma_{vv} = 1$, when the error in x is large.

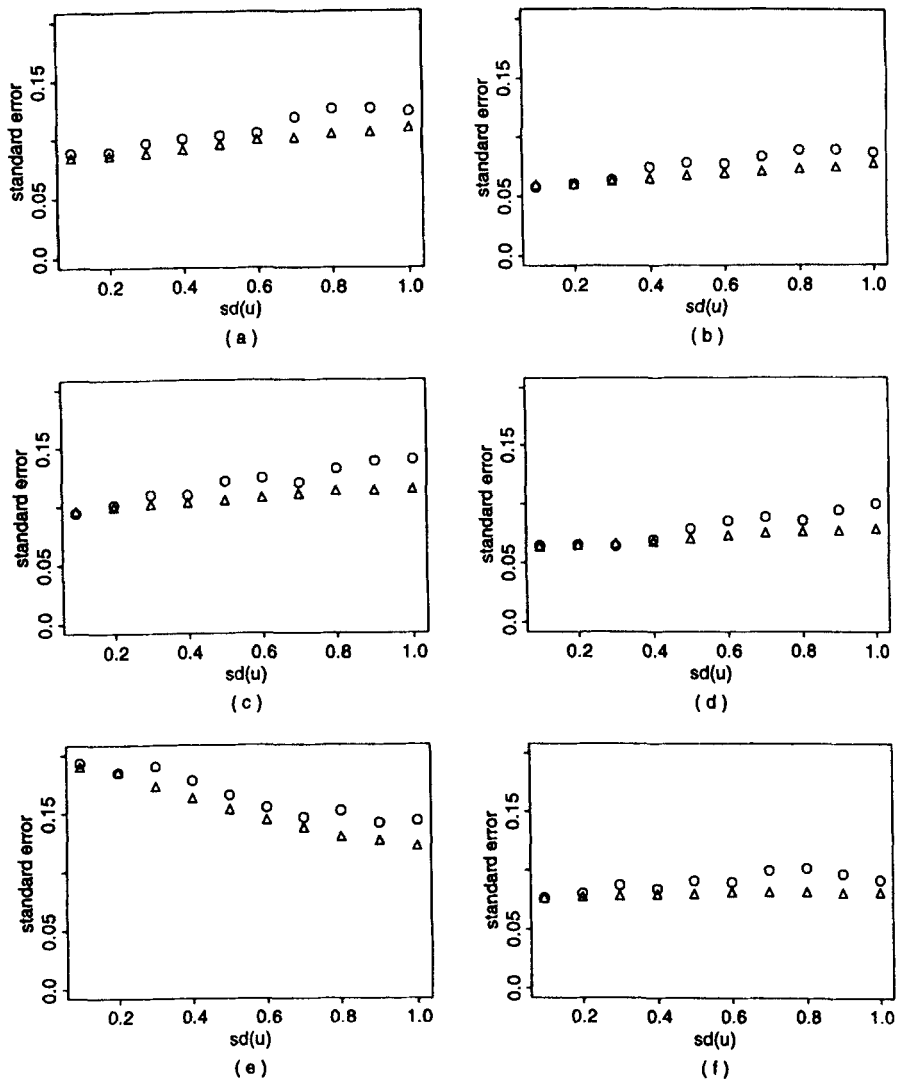

Figure 3. Standard errors (sample O, average asymptotic Δ) for β_2 obtained from 400 non-linear regressions with samples of size 100 under various levels of measurement error and multicollinearity: (*a*) $\rho = 0$, $\sigma(v) = 0$; (*b*) $\rho = 0$, $\sigma(v) = 1$; (*c*) $\rho = 0{\cdot}5$, $\sigma(v) = 0$; (*d*) $\rho = 0{\cdot}5$, $\sigma(v) = 1$; (*e*) $\rho = 0{\cdot}9$, $\sigma(v) = 0$; (*f*) $\rho = 0{\cdot}9$, $\sigma(v) = 1$

An issue which has not been extensively studied in the theory of errors in variables concerns the effect of such errors on the error bands used to assess the significance of various effects. Asymptotic standard errors (SEs) may be used uncritically in the belief that the large sample estimates of these SEs will provide reasonable measures of the true uncertainty about the unknown values of the β's. We investigated the value of these SEs in our simulation study by computing the actual SDs of the simulated β-values. That enabled us to compare these SDs with their theoretical counterparts and see the effects of measurement error and multicollinearity.

Looking at Figures 2 and 3 we see one well-known effect of multicollinearity, the inflation of the SEs, both theoretical and empirical. When $\rho = 0{\cdot}9$ and w is accurately measured, the inflation in the standard errors becomes large for both β-estimators when there is no measurement error. In fact they are almost double their values for $\rho = 0$ (0·2 versus 0·1).

An initially surprising phenomenon occurs when the variables are measured with error and ρ is large: the SEs come down toward their 'correct' levels. Here measurement error has actually favoured the analyst.

We can easily explain this phenomenon on intuitive grounds. Measurement error helps spread out the observed values of X and W. This increases the apparent reliability of the fit. If two points are widely separated, the slope of a line drawn through them will be predictable. The corresponding slope will not be very well-defined when these points are close together.

However, the phenomenon disappears when w is measured with error. Thus multicollinearity impacts on the SE for β_1 only when the surrogate variable is accurately measured. So again, measurement error proves to be the investigator's friend.

In contrast to the situation described above, when ρ is small, measurement error has comparatively little effect. However, Figures 2 and 3 do show that the theoretical SE underestimates the true SEs. While the difference is generally not dramatic, it can become rather large when measurement error becomes large. The investigator relying on estimates of the SE from asymptotic theory could well mistakenly report a significant finding.

More worrying, perhaps, is what our figures do not tell us. Even our extremely simple experiment in non-linear regression demonstrates the rich variety of anomalies which can arise from the interaction of measurement error and multicollinearity. What might happen in a real experiment involving a score of explanatory variables with varying levels of measurement error and multicollinearity?

4. DISCUSSION

We have discussed measurement error in both theoretical and empirical settings. We have described a phenomenon that can arise in regression analysis when explanatory variables are correlated and measured with error. Significance can be transferred from causal explanatory variables to surrogates having only a spurious association with the response variable. Controlling these surrogates would not have the effect on the response expected from controlling its causes.

The authors were motivated to study this phenomenon by the results of an epidemiological investigation. We were looking for the potential association between air pollution and respiratory morbidity and used non-linear regression with longitudinal data.

Our as yet unpublished reanalysis of that presented by Burnett *et al.* (1994) shows that while SO_4 was strongly associated with daily hospital admissions for acute respiratory ailments, SO_2 was not. We knew these two airborne pollutants were correlated. Could 'measurement' error have obscured the real significance of SO_2 and multicollinearity have transferred that significance to SO_4?

Most of the daily and monthly values used in our epidemiological study were interpolated from

a few fixed monitoring sites down to the level of individual census subdivisions in Southern Ontario. Le *et al.* (1994) develop the methodology used. We were attempting to infer the values which would have been measured if an ambient monitoring station had been present. Interpolation error assumed the role of measurement error, given the way the interpolants were used in our investigation.

The accuracy of the interpolator has been empirically assessed using cross-validation studies (Sun 1994 and unpublished work of Le, Sun, Zidek and Burnett). In these studies, the (detrended, deseasonalized and log transformed) data from each monitor was removed, successively, one monitor at a time. The interpolator then used the corresponding data from the remaining monitors to infer the time series of values for the missing monitor.

The values so imputed at all times could be compared with the deleted ('correct') values and the mean squared interpolation errors (MSEs) computed. These MSEs vary from station to station. The location of some stations in relation to the rest makes their values easier to predict. Their MSEs are small.

The MSE over all stations and days turned out to be 0·71 for SO_2 compared with 0·64 for SO_4. The estimated variances for these pollutants are, respectively, 1·38 and 1·35. Thus the error-standard-deviation/standard-deviation ratio for these two transformed pollutants are 0·51 and 0·47, respectively. Using data from ambient monitoring stations, Sun (1994) found the correlations between the daily values of SO_2 and SO_4 to be 0·12. These results, taken in conjunction with the simulation study of the last section, increase our confidence in the significance we attributed to SO_4.

An important measurement error problem arises in epidemiologic air pollution studies when ambient monitors (AMs) are used as surrogates for personal monitors (PMs). As Suh *et al.* (1992) observe

> 'Epidemiologic air pollution studies,..., typically use air pollutant measurements collected from a stationary ambient monitor to reflect exposures for surrounding populations, often for entire communities.'

They conclude that

> '... personal exposure estimates based on ambient concentrations alone may result in substantial misclassification of the exposure status of individuals.'

Information on AM to PM error levels seems fragmentary at best. Measurement errors in PMs themselves would seem negligible, at least for SO_2 and SO_4 among other pollutants (Koutrakis *et al.* 1988; Brauer *et al.* 1989; Suh *et al.* 1992). Also, SO_4 levels observed at a single regional AM predict those at microenvironmental monitors (MMs) with negligible error (Suh *et al.* 1992). So errors entailed in the use of AM measurements of SO_4 to predict PM measurements can be assessed by comparing the MM to PM levels in field studies. Suh *et al.* make this comparison.

From Figure 9 in Suh *et al.* we see that the MM levels of SO_4 are upwardly biased predictors of PM levels. That bias can be removed by regressing the PM levels on the MM levels so as to 'recalibrate' the latter. Suh *et al.* do that regression giving us the recalibrator: $PM = 9{\cdot}03 + 0{\cdot}69\ MM(nmoles/m^3)$. Since $r^2 = 0{\cdot}80$ and the (geometric) standard deviation of

the observed PM levels is $s_{PM} = 1{\cdot}8(\text{nmoles/m}^3)$, we can infer the size of the 'measurement error' as $(1 - r^2)^{1/2} s_{PM} = 0{\cdot}81(\text{nmoles/m}^3)$. We can also infer the standard deviation of the recalibrated MM levels using estimates published by Suh *et al.*: $(0{\cdot}69)s_{MM} = 1{\cdot}31(\text{nmoles/m}^3)$. The ratio $0{\cdot}81/1{\cdot}31 = 0{\cdot}62$ would play the role of $\sigma_{vv} = sd(v)$ introduced in our simulation study.

Undoubtedly, AM to PM measurements of SO_2, a primary pollutant, cannot be assessed through MM levels. However, we have no information on SO_2 related to this issue. Nor have we the correlation between PM levels for SO_2 and SO_4. Thus we cannot fully investigate in this context the impact of the phenomenon addressed in this paper.

We hope this discussion nevertheless illuminates some of the measurement error issues arising in monitoring air pollution. Also, it suggests the need for further study, and for improved design along with methods of analysis. Measurement error working in conjunction with multicollinearity can seriously mislead the analyst into erroneous conclusions.

Our results warn us anew of the importance in epidemiological investigations of careful experimental design with due regard to the management of data quality. No simple after-sampling solution exists for the problem posed by the issues we have identified in this paper, but it cannot be ignored.

ACKNOWLEDGEMENTS

We are indebted to Dr. Michael Brauer for pointing out references to work on measurement error in personal monitoring. Thanks go to Weimin Sun for computing the mean squared interpolation errors quoted in this paper.

REFERENCES

Berkson, J. (1950). 'Are there two regressions?', *Journal of the American Statistical Association*, **54,** 164–180.

Brauer, M., Koutrakis, P., Wolfson, J. M. and Spengler, J. D. (1989). 'Evaluation of the gas collection of an annular denuder system under simulated atmospheric conditions', *Atmospheric Environment*, **23,** 1981–1986.

Burnett, R. T. and Krewski, D. (1994). 'Air pollution effects on hospital admission rates: a random effects modeling approach', *Canadian Journal of Statistics*,

Burnett, R. T., Dales, R. E., Raizenne, M. R., Krewski, D., Summers, P. W., Roberts, G. R., Raad-Young, M., Dann, T. and Brook, J. (1994) 'Effects of low ambient levels of ozone and sulfates on the frequency of respiratory admissions to Ontario hospitals', *Environmental Research*, **65,** 172–194.

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

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.

Koutrakis, P., Wolfson, J. M., Slater, J. L., Brauer, M. and Spengler, J. D. (1988). 'Evaluation of an annular denuder/filter pack system to collect acidic aerosols and gases', *Environmental Science and Technology*, **22,** 1463–1468.

Le, N. D., Sun, W. and Zidek, J. V. (1994). 'Bayesian multivariate interpolation with systematically missing data'. Submitted.

Robinson, W. S. (1950). 'Ecological correlations and the behavior of individuals', *American Sociological Review*, **15,** 351–357.

Simpson, E. H. (1951). 'The interpretation of interaction in contingency tables', *Journal of the Royal Statistical Society, Series B*, **13,** 238–241.

Suh, H. H., Spengler, J. D. and Koutrakis, P. (1982). 'Personal exposures to acid aerosols and ammonia', *Environmental Science and Technology*, **26,** 2507–2517.

Sun, W. (1994). 'Bayesian multivariate interpolation with missing data and its applications', PhD Dissertation, Department of Statistics, University of British Columbia.

Yule, G. V. (1912). 'On methods of measuring association between attributes, *Journal of the Royal Statistical Society, Series B*, **75,** 579–642.

Zidek, J. V. (1984). 'Maximal Simpson-disaggregation of 2 × 2 tables', *Biometrika*, **71,** 187–190.