Statistics in Epidemiology Report

Special Topics Issue!

Measurement Error in Epidemiology

PERSPECTIVES ON MEASUREMENT ERROR

TOR TOSTESON
Dartmouth Medical School

I was very pleased to be asked to contribute to this collection of commentaries on research related to covariate measurement error. This is a very important topic for statisticians and epidemiologists, and it seems like a good time to try to gain perspective on the many

CONTINUED ON PAGE 2

MEASUREMENT ERROR IN EPIDEMIOLOGY — SUGGESTIONS FOR BROADENING THE PERSPECTIVE

MARI PALTA
University of Wisconsin-Madison

Much work in biostatistics has focused on measurement errors in exposure variables and risk factors. The very large literature is summarized in the two books of Carroll et al (1995) [8] and Fuller (1987) [16]. Recent developments included the regression calibration technique by Rosner and Spiegelman [34, 35], the

CONTINUED ON PAGE 7

ESTIMATION AND INFERENCE IN EPIDEMIOLOGY WITH EXPOSURE MEASUREMENT ERROR

DONNA SPIEGELMAN
Harvard University

Since the first papers on this topic appeared in the biomedical research literature [5], much methodologic progress has been made, and there is currently a large and growing number of statisticians working in this area (see Carroll et al (1995) [8] for a recent review). In this

CONTINUED ON PAGE 4

MODEL-BASED LATENT VARIABLE PROBLEMS WITH EPIDEMIOLOGIC APPLICATIONS

ROBERT LYLES
Johns Hopkins University

The potential biases to regression analyses incurred as a result of covariate misclassification or measurement error (ME) are familiar to most epidemiologists, as is the fact that statistical methods are available to adjust for such biases. It is generally appreciated that covariate ME often stems from using less expensive or

CONTINUED ON PAGE 9

TOSTESON, CONTINUED

recent developments in the field. To assist in this effort, I will now share some of my observations and comments. Given space constraints, this discussion will include very few citations, so I must apologize ahead of time to those of you who have contributed to this very active area of biostatistical research for not adequately referencing your work.

I first became interested in measurement error problems in the early 1980's while participating in the Harvard Six Cities Study, a longitudinal study of the health effects of air pollution. Studies conducted by our group had helped to identify the inadequacy of traditional central site monitoring as a measure of personal air pollution exposure. As a result, statisticians and epidemiologists began to search out ways to address the consequences of measurement error in exposures on analyses of the health effects of air pollution. During the same period, similar observations in occupational and nutritional epidemiology lead researchers in other fields to consider the same issues.

The topic of "errors-in-variables" was at that time a venerable subject, with significant references dating back to the 1950's and 1940's. The phenomena of "bias towards the null" in slope estimates for simple regression and the less predictable effects on slope estimates for variables measured without error in multiple regression were well understood and featured in many textbooks. However, most of these techniques were not directly relevant to epidemiolgic regression analyses. In particular, logistic regression, generalized linear models, and proportional hazards models presented new challenges.

By far the most formidable obstacle was, and remains, an inherent computational complexity. These models are characterized by nonlinear regression functions. This feature leads to difficulties in evaluating the proper form for regression function for the expectation of the health outcome variable given the observed, error-prone variable. A good example of this is a simple logistic regression with a normally distributed risk factor and normal measurement errors. The risk model conditional on the error-prone risk factor measure can only be expressed as an integral form. Maximum likelihood estimation thus requires the solution of a indeterminate integral form at each step in the iterative solution of the score equations, a time consuming and numerically unstable procedure.

The recent text by Carroll, Ruppert and Stefanski [8] contains a comprehensive bibliography and summary of the many methods that have been proposed for treating the computational and theoretical issues, including the use of integrable models such as probit regression with normal errors; the regression calibration approximation; small measurement error approximations, which can be viewed as complementing the regression calibration approach; conditional score functions and other methods that attempt to develop improved score equations; approximations based on simulation of measurement error; and improved score tests for basic hyposthesis testing.

The renewed interest among epidemiologists in measurement error methods has also stimulated an interest in design issues. Nutritional and environmental applications required that the classical additive error assumptions for errors-in-variables in linear models be replaced by more general "conditional independence" assumptions for "surrogate" exposure estimates. Also, much of the classical literature assumed that the measurement error was known (or was equal to the residual error in a regression model, as is true in certain engineering applications if measurements are being made with the same instrument). In epidemiology, special validation studies are often required to estimate measurement error parameters. The joint consideration of the design of validation and "main" studies has lead to several interesting methodologic developments, as reviewed by Spiegelman [36].

So where do we stand now, and what directions should we take in the future? There are certainly remain many important methodological issues, such as covariate measurement error in longitudinal data which just now being carefully investigated, with several recent and soon to appear developments. A number of promising new approaches for proportional hazards models have been presented at recent statistical meetings. High grade software for advanced nonlinear techniques is now being developed and distributed, thus removing a practical barrier to the use of the new techniques.

But if I were to identify the most important challenge for future it would to apply measurement error methods in an vital way to important research problems. In preparation for this commentary, I

searched the six year MEDLINE data base for articles in medical journals with any mention of the terms "measurement error" and "regression". This search yielded 22 references. If the search is "restricted to references mentioning correction", "attenuation" , or "adjustment", the number of references drops to three and only two of these apparently involve actual applications of measurement error methods. A similar search on the terms "proportional hazards" and "regression" or "model" yielded 471 references.

It is difficult at best to evaluate the impact of new statistical techniques, but I believe this exercise shows that applications of measurement error methods lag well behind theoretical developments. Why is this? Recently an epidemiologist colleague expressed the view to me that since measurement error adjustments usually have little effect on t-statistics or may even make them smaller, they are regarded as dispensable in studies where nonlinear regression techniques are used primarily to identify "significant" effects. It is hard to counter this argument because, despite constant reminders of the pitfalls of hypothesis testing, many investigators and medical journals remain fixated on the all important "p-value".

There are other reasons one can advance for the scarcity of good applications. I have already mentioned the availability of software, which is currently being addressed. In part, the scarcity of applications may be due to a continuing lack of awareness about the potential problems associated with measurement error, suggesting a need for more review papers in applied journals, workshops, etc. I also think that we have to try harder to use measurement error methods in our collaborative research. My own list of compelling applications includes: meta-analyses in which the included studies have different levels of measurement error; measurement error adjustments for important confounders, since these may have dramatic effects on t-statistics for the risk factor of interest; and sample size adjustments based on estimates of measurement error parameters.

Even without wide application of specific methods, I think research on measurement error methods has made a strong contribution to epidemiology and medical research. It is more common now to see substantial validation studies accompanying large scale studies. This reflects an increasing appreciation of the scientific implications of measurement error. I am optimistic that as these methods are further refined, implemented and disseminated, we will come to view measurement error methods as an essential part of the statistician's tool kit.

★★

1997 YOUNG INVESTIGATOR AWARD
DEADLINE EXTENDED
TO JULY 15, 1997!

The Statistics in Epidemiology Section will once again be awarding a $500.00 prize to be given to the best paper presented at the sessions sponsored by the Section at the annual Joint Statistical Meetings in Anaheim in August 1997. This will be the 3rd year that the section has presented this award. Individuals who are members of the ASA and who are either graduate students of Statistics, Biostatistics or Epidemiology or have graduated at the doctoral or master level within five years of June 1, 1997 are eligible for the award. The paper is to be presented at the meetings by the award candidate.

Individuals who have submitted contributed papers to sessions sponsored by the Section and who wish to be considered for the award in 1997 should send a copy of the paper with a cover letter to:

Susan Ackermann, Ph.D. Past Chair,
Johnson and Johnson,
ICOM Development Group,
920 Rt. 202, P.O. Box 300,
Raritan, New Jersey 08869-0602

by **July 15, 1997.** Please also send a cover letter which includes documentation supporting the eligibility for the award (e.g. certificate of student status or date of graduation). The recipient will be notified by the second week of July 1997. The award and certificate will be presented at the business meeting of the Statistics in Epidemiology Section in Anaheim.

(All candidates should still send a copy of their paper to the Section Program Chair).

★★

SPIEGELMAN, CONTINUED

piece, unless stated otherwise, I will use the terminology *measurement error* to refer to error in both continuous and categorical covariates, while error in the latter is typically referred to in the literature as *misclassification*. This distinction is important because most of the methods proposed for the continuous covariate case are valid for misclassified categorical variables, but the reverse is not true.

Scientific investigators have been slow to apply any of the methodologies thus far proposed to their published work. It has been known for quite some time that random misclassification leads to bias towards the null [23]; hence, estimates of effect could be treated as conservative underestimates of the true underlying effect, an error which has generally been regarded as far less pernicious than overestimation. We know now that this rule has important exceptions, such as when there are more than two categories of exposure [14, 15] and thus the need for methods for removing bias *away* from the null due to measurement error is more important than previously realized.

Unfortunately, there is a dire lack of user-friendly software to implement the methods which have thus far been proposed. Even the *matrix method* [3, 17] for correcting for binary or nominal polytomous categorical covariate misclassification, the oldest and most well known method in epidemiology for dealing with exposure measurement error, has never been implemented with user-friendly software. A notable exception to this is the *regression calibration* method [1, 2, 28, 34, 35]; I and colleagues have written and distributed user-friendly software in two forms (SAS macros, and a Fortran program) to epidemiologists around the world [request via email: stdls@ channing.harvard.edu], and Ray Carroll has also made software available for this method [Obtainable at http://stat.tamu.edu/ under QVF]. This method is useful for cohort studies analyzed by logistic or Cox regression, and for cross-sectional studies using the linear regression model, as long as no recall bias is suspected. It is undoubtedly due, at least in part, to the availability of this software, that this method for correction of exposure measurement error has been applied to analysis appearing in original scientific investigations [32, 42] and many others.

Although many epidemiologists are quite concerned about possible bias in their work due to covariate measurement error, few have the necessary data to incorporate explicit correction for this bias in their analysis. Without augmenting the usual data with a validation or reliability substudy, estimation of the parameters necessary for bias correction is not possible. Statisticians advocate the use of main study/validation study designs when a gold standard exists which can be used to validate the usual method of measurement in a small subsample, and main study/reliability study designs when the error is assumed to be random within-person and no gold standard is available. In this latter design, replicate measures (two is usually sufficient) are taken in a small subsample of the main study participants. Without validation or reliability data, the parameters of the problem are not identifiable unless empirically unverifiable assumptions about the form of the measurement error model as well as the values of its parameters are made. The resulting analysis is extraordinarily sensitive to these assumptions. The main study/validation study has become virtually standard in nutritional epidemiology, but it is rare to find this design used in other branches of epidemiology. Although several papers have provided explicit formulas for sample size and power (see [36] for a review), I am not aware of the availability of any user-friendly software. These calculations involve optimization of complicated multiparameter functions subject to non-linear constraints, and without the availability of such software it is unlikely that epidemiologists will be able to use any methods developed.

I will give a brief overview of the likelihood-based approach. The construction of the likelihood in the presence of covariate measurement error should elucidate the understanding of any approach to this problem, even if the likelihood is ultimately not used. For simplicity, I will give the univariate case -- multivariate extensions are straightforward. If $f_1(Y/x;\beta)$ is the model for the outcome, Y, conditional on covariate x, and β, or some element thereof, is the parameter of interest, and if X is a surrogate for x, i.e. if $f(Y/X,x)=f(Y/x)$, then $f_3(Y/X;\beta,\theta)=\int f_1(Y/x;\beta)\, f_2(x/X;\theta)\, dx$, where $f_2(x/X;\theta)$ is the measurement

error model, indexed by parameter θ. The likelihood in a main study/validation study (M/V) design or a main study/reliability study (M/R) design is thus

$$L(\beta;\theta) = \sum_{i=1}^{n_1} \log[f_1(Y_i \mid X_i;\beta,\theta)] + \sum_{i=n_1+1}^{n_1+n_2} \log[f_2(x_i \mid X_i;\theta)] + \sum_{i=n_1+1}^{n_1+n_2} \log[f_1(Y_i \mid x_i;\beta)]$$

where n_1 is the sample size of the main study and n_2 is the sample size of the validation or reliability study. Standard likelihood estimation and inference can be applied to this model [39].

So why have so many papers been written on methods for consistent estimation and valid inference in models with covariate measurement error? First, f_3 may not have a closed-form solution, requiring the development of mathematical or numeric approximations. For example, with f_1 a logistic regression model and f_2 a Gaussian model, f_3 has no closed form. This important case motivated many recent papers, as it describes a group of models typically used in epidemiologic applications [7, 13, 41].

Second, although f_2 and f_1 are typically standard models supported by commercial statistical software, f_3 rarely is, even when it is closed-form. In addition, equation (1) is almost never supported by commercial software. $L(\beta,\theta)$ given in equation (1) is typically a complicated, non-linear multi-parameter function, often with linear and/or non-linear constraints on elements of (β,θ). Estimation and inference using (1) must be customized to the problem at hand, and the assistance of a technically sophisticated Fortran or C+ programmer is thus required.

Third, this is a fully parametric method, and there is concern that bias in the estimation of β might be introduced by mis-specifying f_2. In the estimating equations tradition, methods have been proposed where the first (and possibly second) moments of f_1 are specified. Some authors specify the first (and possibly second) moment(s) of f_2 and others the full model for f_2 [9, 19]. In either case, estimating equations for β are then derived, and estimation and inference proceeds as usual. In much of this work, in my view, inadequate attention is paid to explicit modeling of the measurement error process or the additional uncertainty in the inference due to uncertainty about the measurement error process, given the data. In a third approach, the $E(Y/x;\beta)$ is specified, and estimating equations are derived with f_2 completely non-parametric [31, 33]. Even when the estimating equations are semi-parametrically efficient, in examples I have looked at, there is a huge loss of efficiency relative to the fully parametric likelihood-based method [38]. However, this method is non-parametric in exactly the way that is needed -- in the specification of f_2, about which there is typically a modest amount of data.

In conclusion, much interesting and practical work has been published by statisticians on methods for valid estimation and inference in the presence of exposure measurement error. Some areas which could benefit from further attention, in my view, include software development for both design and analysis, research into the effect of local mis-specification of the relevant features of the measurement error model, methods for failure-time data, and more efficient, intuitive methods for misclassified data and data with both continuous variable measurement error and categorical variable misclassification.

★★

EPIDEMIOLOGY SECTION SPONSORED SESSIONS AT JSM IN ANAHEIM

INVITED SESSIONS:

T 10:30: Issues in Health Policy
T 2:00: SES and Health in Minority Populations
W 8:30: Meta Analysis in the New Millennium
W 2:00: Environmental Epidemiology
Th 8:30: Design Issues in Genetic Epidemiology

CONTRIBUTED SESSIONS:

M 8:30: Applications of Survival Analysis
M 2:00: Analysis of Time Series Data
T 8:30: Appl. of Stat. Models in Epidem I

T 2:00: Appl. of Stat. Methods in Epidem I
W 8:30: Appl. of Stat. Models in Epidem II
Th 2:00: Appl. of Stat. Methods in Epidem II

W 12:00: Contributed Poster Session

PALTA, CONTINUED

bootstrap extrapolation of Stefanski and the introduction into biostatistics of instrumental variable techniques. These techniques tend to boost the estimate of association between the risk factor and outcome, but as they also increase the standard error, leave significance levels relatively unchanged. Finding examples of the application of these techniques in epidemiology, remains a challenge. Looking through the Methods sections of all articles in 7 recent issues of the American Journal of Epidemiology, I found one article (besides [37]) that used any sort of correction for measurement error or within individual variability. In that article, which had repeat within individual measurements on blood pressure and urinary sodium:potassium ratio, the authors stated "Although a correction factor was estimated, the primary emphasis for these analyses remained on the raw values". Interestingly, the authors found that their correlation coefficient between urinary sodium:potassium ratio and blood pressure increased from 0.16 to 0.28 after correction for "regression dilution", a substantial difference which was primarily due to variability in urinary values [18]. One wonders at the shyness in utilizing error adjustments, especially as many risk factors, including most of those analyzed in the articles I scanned, are measured with error. I propose as possible reasons the lack of access to software, reluctance to oversell risk factors, complexity of adjustment when multiple risk factors and confounders are involved, and concern with assumptions. I also point to some aspects of measurement error modeling in epidemiology that have received relatively less interest from biostatisticians.

Two recent publications are examples of the type needed to bridge some of the gap by making assumptions and procedures very clear. The technique of instrumental variables, was presented by Carroll and Stefanski [10]. Another very recent publication [37] shows when the regression calibration technique can be applied in the situation without a true gold standard, and in the process discusses how to estimate the correlation between errors using a third measurement. By discussing the different assumptions involved, these developments may re-kindle interest among a broader epidemiologic audience. Although the effect of various independence and distributional assumptions has been investigated in the statistics literature, not much of this has reached the epidemiologic community.

Another road to increase interest in measurement error adjustment among epidemiologists may be via the deep concern with confounding. In most observational studies, the fear of falsely concluding association is paramount. However, adjusting for measurement error in multiple risk factors and confounders is complex, both statistically and logistically, as a multivariate validation sample must be assembled [34]. At this point such validation samples are not routine. With confounder error following the calibration model, so called "resonant" error occurs in the estimated association between the outcome and risk factor. In this situation it is unpredictable how the risk factor-outcome association will be affected. A recent article [22] showed that the consequences of confounder error in strong confounders may be large. They suggested that associations in observational studies between dietary beta-carotene and lung cancer may have been due to resonance of error in estimating smoking exposure [22]. Clearly such possibilities should be of great concern to epidemiologists and biostatisticans alike.

As in modeling measurement error in the risk factor, there is a choice of whether the confounder error is assumed to follow a regression calibration or a Berkson model [8]. The first model assumes independence between the true value and the error ($W=X+U$, where W, X, U are observed, true and error values), while the second model (in its basic form) assumes independence between the observed value and the error ($X=W+U$). While the calibration error structure seems natural when the confounder is known, but mismeasured, it may be argued that the latter may describe some situations where not quite the right confounder has been found (e.g. when using mother's education to capture socioeconomic level). When formulating a Berkson error structure for the confounder, one finds that measurement error will

bias ordinary regression only if the distribution of U depends on the risk factor. It is not known, however, how often and how strongly this happens. In a recent publication, we outlined how Berkson error can be transformed into confounding [26]. Some of our work with confounding in longitudinal studies is also related to the above, as it shows that one can adjust for cross-sectional confounders (or cohort effects) by using the individual specific risk factor mean as a surrogate for the confounders. This produces an unbiased estimator of risk factor effect as long as time to time variability in the confounder is not correlated with time to time variability in the risk factor [11]. We discussed a study of sleep disorders across two time points, four years apart. The logistic regression coefficient of "feeling excessive daytime sleepiness" (dichotomized into "never or rarely" versus "sometimes") on age was -0.15 ($p<.0001$) in GEE analysis, but changed to 0.036 ($p=.0003$) when mean age of each individual across surveys was adjusted for. The investigators saw the mean age as a surrogate for life style factors which may cause sleep deprivation in younger persons, thereby causing a seeming decrease in sleepiness with aging.

In model formulation, Berkson confounder error is indistinguishable from Berkson error in the risk factor, and also from error in the response. Carroll et al [8] pointed out that the likelihood for the Berkson model is identical to that of a mixed effects model. Subject matter knowledge, of course, may identify the errors that are operating in a specific situation. As we know, measurement error in the response creates no bias in regression coefficients in ordinary regression, only lack of statistical efficiency. However, in the grouped and dichotomized responses often encountered in epidemiology the case is different. Classical epidemiologic (and biostatistical) texts (e.g. Breslow and Day [4]) have dealt with the problem in terms of "percent mis-classification" and have indeed shown that, given certain independence conditions, there may be a conservative bias in the estimated odds ratio. Social scientist appear to have a more elegant formulation in terms of "latent variables" (see e.g. [24]). It can be argued that many ordinal and binary responses arise from underlying continuous scales. As this scale is (almost) imaginary it might as well be assumed scaled to be normally (or logistically) distributed, hence leading to probit and logit links. This is very helpful, as it allows us to fall back on the usual variance component framework from multivariate theory to model binary and ordinal outcomes.

Work by [25] and Qu et al. [29] is relevant in modeling the correlation structure, including error, in the latent outcome. As the correlation structure is modeled, application in longitudinal settings is natural. For example, in the study of sleep disorders described above, we asked several questions related to tiredness. The responses were on an ordinal scale from 0-5, but were dichotomized in analyses. The structural latent variable framework [27] allowed the following variance components to be modeled (a) between individual variation in intercept, which is equivalent to correlation within individual (b) variation between two tiredness questions on the questionnaire at the same time point (c) correlation in the error of the same item across the two time points. However, in modeling latent variables, some aspect of the variance has to be assumed as equaling 1, i.e. chosen as the unit of regression coefficients. Typical choices are total variance (= marginal models) and within individual variance (=cluster specific models). When the error- in -response variance is estimated it can be removed from the standardizing variance, which amounts to rescaling to the variability of the true outcome. Regression coefficients are thereby adjusted for measurement error. In our example, with marginal coefficients this resulted in an increase of approximately 20% in regression coefficients of tiredness on age, gender and BMI. As epidemiologists turn to global health assessment outcomes measured by questionnaires, such assessment will become important.

In summary, then, much work remains in increasing understanding and application of measurement error techniques in epidemiology. It appears necessary to acknowledge greater complexity, while retaining interpretability. Methods from the social sciences, where there is a stronger tradition of assessing measurement errors, may provide meaningful insights. We also need to pay careful attention to the concerns of epidemiologists, who are aware of what assumptions do and do not hold in their data, primarily worry about confounders and are professionally geared to conservatism with weak associations.

★★

more readily obtainable surrogate exposure variables. It may be less common to contemplate the ME problem that arises when one is forced to characterize an important aspect of individuals' exposure experience on the basis of a small collection of actual exposure data. For example, the true mean level of some exposure experienced over a study period may be biologically relevant to an outcome of interest, but one may only have a few repeated measurements available on each individual from which to calculate a sample mean as a surrogate.

If a reasonable model for repeated exposure measurements can be postulated, the ME resulting from the inability to observe the actual unknown exposure variable (in this case the latent "true" mean) can be adjusted for using existing techniques. These techniques are in some cases identical to, and are at least the same in spirit as, the method commonly termed "regression calibration". This method has well-demonstrated epidemiologic applications in the contexts of linear, logistic, and proportional hazards regression models [40].

Consider the case of a longitudinal study in which subjects are measured periodically with respect to an exposure variable (e.g., a biomarker, a shift-long personal breathing zone measurement, etc.). Assume for simplicity that the distribution of repeated exposure measurements is stationary with a compound symmetric correlation structure, and that *n* measurements are obtained on each of *k* subjects. Then a reasonable overall model might be a simple random effects ANOVA, e.g.,

$$Y_{ij} = \mu + b_i + e_{ij} \ , \ i = 1,\ldots,k \ ; \quad j = 1,\ldots,n$$

where we assume

$$b_i \sim N(0, \sigma_b^2), \text{ and}$$

$$e_{ij} \sim N(0, \sigma_w^2)$$

are independent. A latent variable ME problem can then be illustrated in terms of a "true disease model" (TDM), a "measurement error model" (MEM), and a "predictor distribution model" (PDM) [12]:

$$TDM \ R_i = \alpha + \beta\mu_i + \varepsilon_i \quad , \varepsilon_i \sim N(0, \sigma^2)$$

$$MEM \quad : \quad \overline{Y}_i = \mu_i + \overline{e}_i$$

$$PDM \ \mu_i = (\mu + b_i) \sim iid \ N(\mu, \sigma_b^2) \ .$$

The TDM is a simple linear regression of a continuous health-related outcome (R) on true individual-specific mean exposure (μ_i). The MEM relates the sample mean of the *n* observations on the i-th person (which is the logical surrogate exposure variable) to the true mean, via the MEM term $\overline{e}_i$ Note that MEM and PDM are directly specifiable based on the overall model for repeated exposures. Hence, we have a straightforward additive-normal ME problem with non-systematic and non-differential error. It is easily shown that the OLS estimator from the surrogate regression of R_i on $\overline{Y}_i$ has the convergence property

$$\hat{\beta}_{OLS} \rightarrow (1 + \lambda/n)^{-1} \beta \quad ,$$

where $\lambda = \sigma_w^2 / \sigma_b^2$ [6,30]. Regression calibration provides the corrected estimator $\hat{\beta}_{RC} = \hat{\beta}_{OLS} (1 + \hat{\lambda}/n)$, where $\hat{\lambda}$ is a consistent estimator (such as the MLE) for λ. An adjusted confidence interval may also be derived via application of the delta method [40]. The adjusted estimator $\hat{\beta}_{RC}$ is essentially unbiased and the coverage of the adjusted confidence interval (which accounts for estimation of λ) can be shown via simulation to maintain the nominal value of 95% in all cases. The biased naive confidence interval, however, will not contain the true value for large sample sizes.

The example above simply couches the well-known additive-normal ME scenario in simple linear regression within a particular model-based construct. However, the epidemiologic appeal of this construct (or primarily its extensions) may be considerable for situations in which biological arguments point to average exposure over a study period as an important risk factor for an outcome. These arguments are more appropriately embodied in a conceptual health effects model like the TDM model above than in the naive regression model employing a potentially crude surrogate for true average exposure. The overall model for repeated exposure measurements, once validated, provides the mechanism by which the appropriate study question can be addressed. A crucial general point is that the appeal of adjustment procedures like regression calibration is in their preservation of <u>validity</u> with respect to point and confidence interval estimation in the presence of

measurement error. For instance, if one were solely interested in testing whether $\beta = 0$ in the above setting, one would do no better than the usual t-test based on the surrogate regression model. For insight into the incorporation of additional covariates measured with or without error in the TDM, one can refer to the vast literature on ME adjustment.

While it is possible to imagine many practical problems in which the basic conceptual framework illustrated here might be applicable, most of these will not permit straightforward regression calibration as a solution. If the exposure data are unbalanced, for instance, then it is not possible to derive a simple convergence result analogous to the above. In these cases one must appeal to more general methods, although the basic goal and conceptualization of these methods is essentially the same. One possible approach would be an adaptation of quasi-likelihood [19], which incorporates heterogeneity in the conditional variance of the response variable given the surrogate. This strategy has been useful in an occupational epidemiologic application involving grouped data on workers, with the predictor of interest being the true mean exposure over a study period based on a model assuming lognormality of repeated exposure measurements [20]. Adaptations

of the conceptual framework to address further epidemiologic hypotheses of potential interest are also possible. In this direction, as part of an ongoing study of HIV infection, an investigation of the health-related effects of both individual-specific mean and variability with respect to biomarker levels over time has been undertaken using a natural extension of the setting illustrated via models decribed above [21].

[Dr. Lyles is the Winner of the 1996 Epidemiology Section Young Investigator Award]

REFERENCES

[1] Armstrong B. *Comm Stat Sim Comp* 1985; 14:529.
[2] Armstrong B, Whittemore AS, et al. *Stat Med* 1989; 8:1151.
[3] Barron BA. *Biometrics* 1977; 33:414.
[4] Breslow NE, Day NE. Statistical Methods in Cancer Research, Volume 1, IARC No. 32, Lyon, 1980.
[5] Bross I. *Biometrics* 1954; 10:478.
[6] Brunekreef B, et al. *Am J Epid* 1987; 125:892.
[7] Carrol RJ, et al. *Biometrika* 1984; 71:19.
[8] Carroll RJ, Ruppert D, Stefanski LA. Measurement Error in Nonlinear Models. Chapman & Hall 1995.
[9] Carroll RJ, Stefanski LA. *JASA* 1990; 85:652.
[10] Carroll RJ, Stefanski LA. *Stat Med* 1994; 13:1265.
[11] Chao W, Palta M, et al. *Biometrics* 1997; 53:47.
[12] Clayton D (1992). In: *Statistical Models for Longitudinal Studies of Health*, Oxford Univ. Press, U.K.
[13] Crouch EAC, Spiegelman D. *JASA* 1990; 85:464.
[14] Domemeci M, Wacholder S, et al. *Am J Epid* 1990; 132:746.
[15] Flegal KM, Keyl PM, et al. *Am J Epid* 1991; 134:1233.
[16] Fuller WA. Measurement Error Models. John Wiley & Sons 1987.
[17] Greenland S. *Stat Med* 1988; 7:745.
[18] Kaufman JS, Owoaje EE, et al. *Am J Epi* 1996; 143:1203.
[19] Liang K and Liu X (1991). In: *Estimating Functions*, Oxford Univ. Press,U.K.
[20] Lyles R and Kupper L. "A detailed evaluation of adjusment methods for multiplicative measurement error in multiple linear regression, with applications in occupational epidemiology." *Biometrics* (in press).
[21] Lyles R, Muñoz A, et al. "A measurement error-adjusted analysis with latent biomarker variability as a covariate. Submitted.
[22] Marshall JR, Hastrup JL. *Am J Epi* 1996; 43:1069.
[23] Mote VL, Anderson RL. *Biometrika* 1965; 52:95.
[24] Muthen BO. *JASA* 1979; 74:807.
[25] Muthen BO. LISCOMP- Analysis of linear structural equations with a comprehensive measurement model. Mooresville, IN: Scientific Software. Inc, 1988.
[26] Palta M, Lin C-Y, et al. "Effect of confounding and other misspecification in models for longitudinal data In: Modeling Longitudinal and Spatially Correlated Data." Lecture Notes in Statistics Series, 122. Proceeding of the Nantucket Conference on Longitudinal and Correlated Data. Springer-Verlag 1997.
[27] Palta M, Lin C-Y. "Latent variables, measurement error and methods for analyzing longitudinal binary and ordinal data." Technical Report #106, Department of Biostatistics, University of Wisconsin, Madison, 1996.
[28] Prentice RL. *Biometrika* 1982; 69:331.
[29] Qu Y, Williams GW, et al. *Biometrics* 1992; 48:1095.
[30] Rappaport S, Symanski E, et al *Environ Health Persp*, 1995; 103:49.
[31] Reilly M, Pepe MS. *Biometrika* 1995; 82:299.
[32] Rimm EB, et al. *JAMA* 1996; 225:447.
[33] Robins JM, Hsieh F, et al *JRSSB* 1995; 57:409.
[34] Rosner B, Spiegelman D, et al. *Am J Epi* 1990;132:734.
[35] Rosner B, Spiegelman D, et al. *Am J Epi* 1992; 136:1400.
[36] Spiegelman D. *J Stat Plan Inf* 1994; 42:187.
[37] Spiegelman D, Scneeweiss S, 1997. *Am J Epi* 145:184.
[38] Spiegelman D, Casella M. Fully parametric and semi-parametric regression models for common events with covariate measurement error in main study/validation study designs. Biometrics 1997, in press.
[39] Spiegelman D, Rosner B. "Estimation and inference for logistic regression with covariate misclassification and measurement error in main

study/validation study designs" Submitted manuscript, 1997.
[40] Spiegelman D, McDermott A and Rosner B. *Am J Clin Nutr* (in press).
[41] Stefanski LA and Carroll RJ. *Ann Stat* 1985; 13-1335.
[42] Willett WC, Hunter DJ, et al *JAMA* 1992; 268:2037.

★★

1997 SECTION OFFICERS

Section Chair
Marlene J. Egger, Ph.D.
Division of Epidemiology, Biostatistics and Prevention Research
University of Utah School of Medicine
801-581-7234 (-2759 FAX)
megger@dfpm.utah.edu

Section Chair Elect
Gladys Reynolds, Ph.D.
Centers for Disease Control Office of Minority Health
1600 Clifton Road, MS-A50
Atlanta, GA 30333
404-639-7224 (-7039 FAX)
ghr1@odod1.em.cdc.gov

Past Section Chair
Susan P. Ackermann, Ph.D.
Cephalon, Inc.
610-344-0200 (-0649 FAX)
sackerma.cephalon@notes.ccmail.compuserve.com

Section Secretary/Treasurer
Robert D. Abbott, Ph.D.
Division of Biostatistics
University of Virginia School of Medicine
804-924-1687
rda3e@kronos.pharm.virginia.edu

Program Chair
Glen David Williamson, Ph.D.
Div of Surveillance and Epidemiology
Centers for Disease Control
404-639-3806 (-1546 FAX)
dxw2@epo.em.cdc.gov

Program Chair Elect
Ralph D'Agastino, Jr., Ph.D.
Department of Public Health Services
Bowman Gray School of Medicine
Winston-Salem, NC 27157
910-716-9410 (-5425 FAX)

Section Representative
Deborah V. Dawson, Ph.D.
Dept of Epidemiology and Biostatistics
Case Western Reserve University
216-778-8476 (-8297 FAX)
dvd2@po.cwru.edu

Publications Officer
Lance A. Waller, Ph.D.
Division of Biostatistics
University of Minnesota School of Public Health
612-625-2138 (-60660 FAX)
lance@muskie.biostat.umn.edu

STATISTICS IN EPIDEMIOLOGY REPORT
American Statistical Association
1429 Duke Street
Alexandria, VA 22314-3402
USA

Statistics in Epidemiology Report is the official publication of the Section on Statistics in Epidemiology of the American Statistical Association. All communications regarding this publication should be forwarded to: Stephen J. Gange, Ph.D., Editor, *Statistics in Epidemiology Report*, Department of Epidemiology, Johns Hopkins School of Hygiene and Public Health, 615 N. Wolfe St -- Room E7010, Baltimore, MD 21205, 410-955-4320 (-7587 FAX), **sgange@statepi.sph.jhu.edu**

All communications regarding membership in the ASA and the Section on Epidemiology, including change of address, should be sent to: American Statistical Association 1429 Duke Street, Alexandria, VA 22314-3402

NONPROFIT ORG.
U.S. POSTAGE
PAID
BALTIMORE, MD
PERMIT NO. 5143