Correcting for Bias in Relative Risk Estimates Due to Exposure Measurement Error: A Case Study of Occupational Exposure to Antineoplastics in Pharmacists

Donna Spiegelman, ScD, and Barbara Valanis, DrPH, FAAN

ABSTRACT

Objectives. This paper describes 2 statistical methods designed to correct for bias from exposure measurement error in point and interval estimates of relative risk.

Methods. The first method takes the usual point and interval estimates of the log relative risk obtained from logistic regression and corrects them for nondifferential measurement error using an exposure measurement error model estimated from validation data. The second, likelihood-based method fits an arbitrary measurement error model suitable for the data at hand and then derives the model for the outcome of interest.

Results. Data from Valanis and colleagues' study of the health effects of antineoplastics exposure among hospital pharmacists were used to estimate the prevalence ratio of fever in the previous 3 months from this exposure. For an interdecile increase in weekly number of drugs mixed, the prevalence ratio, adjusted for confounding, changed from 1.06 to 1.17 (95% confidence interval [CI] = 1.04, 1.26) after correction for exposure measurement error.

Conclusions. Exposure measurement error is often an important source of bias in public health research. Methods are available to correct such biases. (*Am J Public Health.* 1998;88:406–412)

Introduction

Measurement error and misclassification of exposure are widely acknowledged as pervasive and often important sources of bias in much public health research.[1,2] Measurement error in variables of interest contributes critically to the difficulty in answering certain scientific questions with major public health implications, such as in studies of the long-term health effects of high-fat diets[3] and exposure to electromagnetic fields[4] and radon gas.[5] There has been an abundance of interest in the topic of covariate measurement error and misclassification in recent years, and the reader is referred to recent nontechnical reviews of this literature.[6–9]

The biases due to measurement error and misclassification have been known for quite some time. One of the earliest papers on this topic in the biomedical literature appeared in the Journal in 1962,[10] using an example of bias in estimating the relationship between cervical cancer and circumcision due to misclassification of this exposure variable. Despite the plethora of work investigating such biases and the extensive development of methods to correct for them following the publication of early seminal works,[10–12] very few original scientific publications have used these methods.[13,14] Investigators can no longer rely on the maxim that as long as measurement error or misclassification is nondifferential, estimates of effect will be, at worst, conservative underestimates of the true underlying effect, since recent research has uncovered circumstances under which this rule fails to hold.[15–18]

This paper compares 2 approaches to point and interval estimation of risk ratios for binary data when 1 or more exposure variables have been measured with error and/or misclassified: a completely general likelihood-based approach and a conceptually and computationally simpler method that makes more restrictive assumptions (regression calibration). We focus on obtaining relative risk point estimates that are not contaminated by bias due to covariate measurement error as well as on obtaining efficient, correctly centered confidence intervals (CIs) that have the specified coverage probability. Spiegelman and Casella[19] further developed the statistical methodology discussed.

The methodology is formulated for main study/validation study designs, in which data obtained in a subsample measure exposure(s) without error using some superior and costly technology, thus permitting explicit modeling of the covariate measurement error/misclassification process. The validation data permit unbiased estimation of the parameters of interest without invoking empirically unverifiable assumptions about measurement errors. In addition, the validation data permit correct representation of the true uncertainty in the resulting inference using confidence intervals that reflect all sources of variability in the data, including those due to covariate measurement error.

This paper assumes that other sources of bias to which public health data are vulnerable (e.g., bias resulting from unmea-

Donna Spiegelman is with the Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, Mass. Barbara Valanis is with the Kaiser Permanente Center for Health Research, Portland, Ore.

Requests for reprints should be sent to Donna Spiegelman, ScD, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Ave, Boston, MA 02115 (e-mail address: stdls@channing.harvard.edu).

This paper was accepted May 30, 1997.

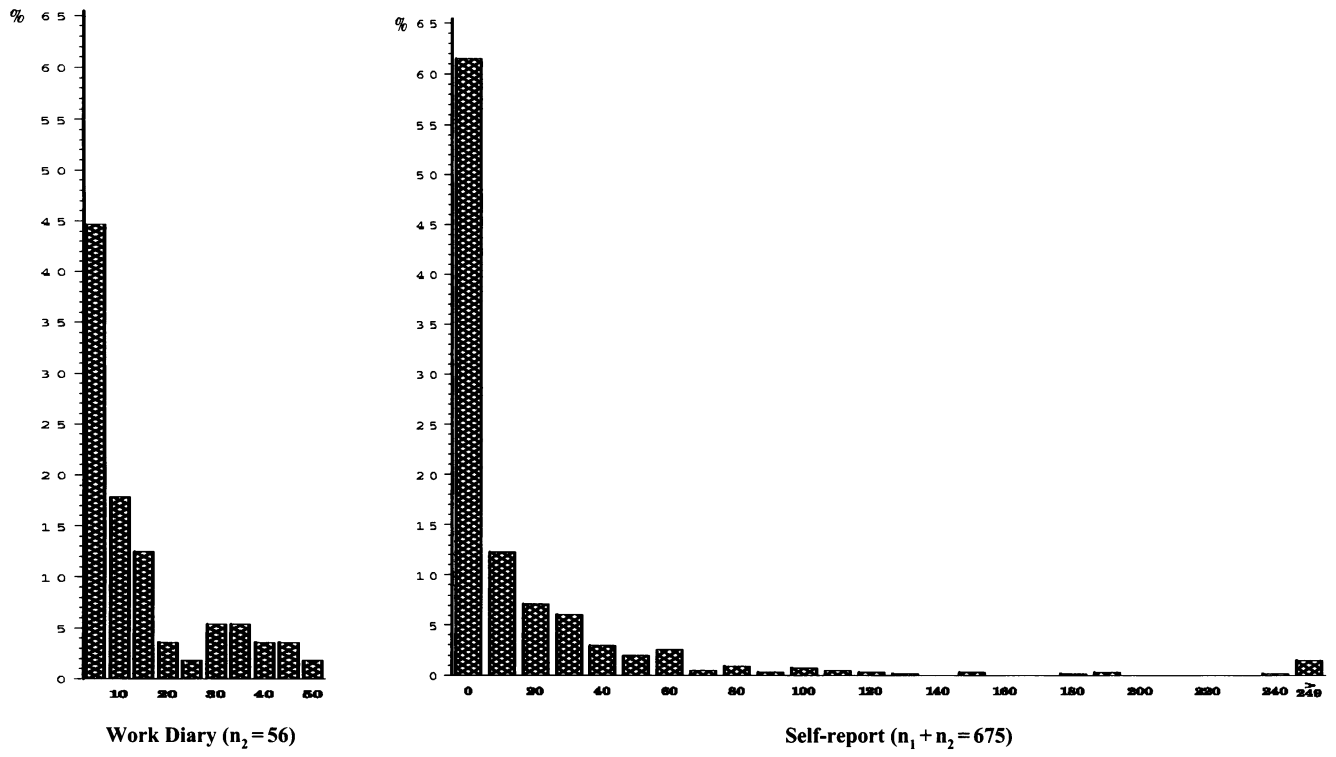

FIGURE 1—Distribution of drugs mixed per week.

sured confounding or from improper selection of the study population) have been eliminated through proper study design and analysis. We illustrate these methods by developing a detailed example from Valanis and colleagues' 1993 main study/validation study of the acute health effects of chemotherapeutics among pharmacists and pharmacists' aides.[20]

Materials

Valanis and colleagues reported a multicenter study of the relationship between occupational exposure to antineoplastic drugs and the prevalence of acute health effects among pharmacists and pharmacy technicians[20] and among nurses and nurses' aides[21] (the Health and Occupational Exposure to Anti-Cancer Drugs Study). We analyzed data on 675 pharmacists and pharmacy technicians from this study. Data relating exposure to the prevalence of 27 symptoms experienced during the previous 3 months were collected through self-administered questionnaires in both the first and second phases of the study. In this analysis, we defined the exposure variable of interest as average weekly number of antineoplastic drugs mixed. This exposure was self-reported on the questionnaire by all study participants. In the second study phase, the validity of this exposure variable was assessed among 56 (8%) of the 675 participants by comparing average weekly number of antineoplastic drugs mixed, as self-reported on the questionnaire, with data calculated from 1- to 2-week on-site diaries of drug mixing activity. Data were available on symptom prevalence, and other possible determinants of prevalence included sex, age, type of workplace, type of shift, educational level, work stress, and current cigarette smoking status in both study phases. Except for exposure, these variables were assumed to be known without error. Further details on the design of this study and its data collection procedures have been published previously.[20,22]

The current investigation was not intended to be a definitive analysis of the effects of occupational antineoplastics exposure on acute health effects; rather, it was intended to demonstrate the feasibility and utility of statistical methods that eliminate bias in point and interval estimates of effect due to exposure measurement error in a realistic example of some public health importance. Fever was selected because the prevalence of fever was high in the study population, suggesting sufficient statistical power for undertaking the analysis. An increased prevalence of fever could reflect infection resulting from the lowering of white blood cell count by exposure to antineoplastics. Low white blood cell counts can increase susceptibility to infection, which in turn can produce fever. One of the Health and Occupational Exposure to Anti-Cancer Drugs Study pilots showed some white count differences between exposed and control groups.[23]

Fever prevalences were 17% in the main study and 11% in the validation study. The average weekly numbers of antineoplastics mixed, according to questionnaire reports, were 21.3 (SD = 103.7, range = 0 to 1666) and 39.1 (SD = 24.5, range = 9 to 117) in the main study and the validation study, respectively. According to diary reports, the average weekly number of antineoplastics mixed was 13.8 (SD = 24.5, range = 1 to 50) in the validation study. The main study and validation study had similar distributions of the potential confounders (sex, age, smoking status, educational level, primary place of employment, and shift worked). Figure 1 presents the distribution of the exposure, as assessed in the main

study through self-reported questionnaire and as assessed in the validation study through 1- to 2-week work diaries. As is typical of many occupational and environmental exposures, the distribution in both cases was sharply skewed. Figure 2 is a scatter plot of the 2 methods of exposure assessment in the validation data. The correlation coefficient between the 2 methods was 0.70. These data show a moderate amount of exposure measurement error, which is most likely to lead to bias in the point estimate of effect and underestimation of the variability inherent in the data, as reflected by the usual interval estimate.

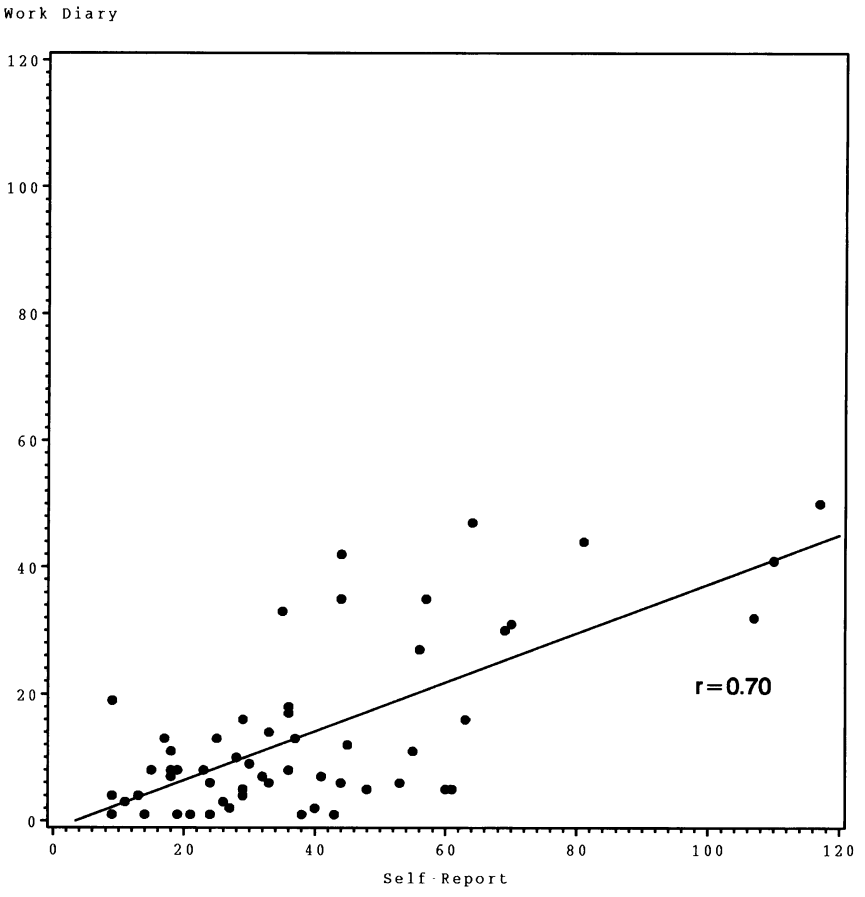

FIGURE 2—Number of drugs mixed per week: work diary vs self-report ($n_2 = 56$).

Regression Calibration

In this method,[24,25] the point and interval estimates of effect are first obtained by fitting the following logistic regression model:

$$(1) \quad \text{logit}\,[Pr\,(D = 1|X)] = \beta_0 + \beta X,$$

where X is the covariate measured with error and β is the log odds ratio for a 1-unit increase in X. When measurement error is present in X, the estimated regression coefficient, $\hat{\beta}$, is biased—often severely so—relative to what it would be on the basis of the correct model, logit $[Pr\,(D = 1|x)]$, where x is measured without error. In this model, $D = 1$ may indicate the cross-sectional presence of a symptom, in which case the logistic regression coefficients estimate the log prevalence odds ratio. In a prospective cohort study, $D = 1$ may indicate the occurrence of disease by the end of a follow-up period of duration t. When the disease is rare, the log risk (or prevalence) odds ratio closely approximates the log risk (or prevalence) ratio.

The estimated logistic regression coefficients from equation 1 can then be adjusted for bias due to measurement error in a simple 1-step procedure. When there is 1 covariate in the relative risk model and this covariate is measured with error, Rosner et al.[24] proposed that the point and interval estimates of log relative risk can be corrected for measurement error by means of the following formulas:

$$\hat{\beta}_{RC} = \frac{\hat{\beta}}{\hat{\gamma}}$$

and

$$\hat{Var}\,(\hat{\beta}_{RC}) = \frac{1}{\hat{\gamma}^2}\,\hat{Var}\,(\hat{\beta}) + \frac{\hat{\beta}^2}{\hat{\gamma}^4}\,\hat{Var}\,(\hat{\gamma}),$$

where $\hat{\beta}$ is estimated from equation 1, $\hat{\beta}_{RC}$ is the corrected logistic regression coefficient, and $\hat{\gamma}$ is obtained from fitting the linear regression model given by

$$(2) \quad x = \alpha' + \gamma X + \epsilon,$$

where x is the correctly measured exposure variable and ϵ is a random error term with mean 0. For the more realistic case when there is more than one covariate in the relative risk model (measured with error or not), these formulas were subsequently generalized.[25] The regression calibration method has been discussed by other authors[26] and shown to be applicable to probit regression problems,[27] survival data models,[28] and normal discriminant models.[29] (Fast, user-friendly SAS macros [The SAS Istitute, Inc, Cary, NC] and Fortran programs that perform these calculations in the main study/validation study design can be obtained from D. S.) An analogous procedure is used when continuous covariates are measured with random within-person variability and no gold standard is available (e.g., in the case of blood pressure and serum cholesterol).[30] For this situation, repeated measurements of the biological variable are needed only in a subsample, the reliability substudy.

The corrected logistic regression coefficient, $\hat{\beta}_{RC}$, was derived under the assumptions of a rare disease, small relative risk, small measurement error, and a measurement error process that is independent of disease status. Its variance estimate further assumes that the residual variance from the linear measurement error model (equation 2) is constant across levels of the data. Rosner et al. explored, through simulations in the univariate case, the limits of violations of these assumptions.[24] The odds ratio obtained from the logistic model may not be the parameter of interest in public health research,[31] and the accuracy of its approximation to the corresponding risk or prevalence ratio decreases as disease frequency increases and as the risk or prevalence ratio increases. The assumptions required for

valid use of $\hat{\beta}_{RC}$ for prevalence ratio estimation may not be realistic in many situations, such as in Valanis and colleagues' study,[20] in which the disease of interest had a prevalence of 17% and the error variance appeared to increase with increasing level of drugs mixed. Thus, we present a completely general maximum-likelihood–based procedure that makes no a priori assumptions about the prevalence of the outcome or the measurement error/misclassification process, and we compare results from an example of this procedure with results using $\hat{\beta}_{RC}$.

The Maximum-Likelihood Approach

Here the data analyst has the freedom to develop the model that is best suited to the research goals and the data at hand.[19] This method can be applied when 1 or more continuous determinants of outcome are measured with error and, in addition, when 1 or more categorical determinants are misclassified. The investigator first specifies a relative risk model, $f_1(D|x,U_1;\beta)$, for which the parameter(s) of interest will be 1 or more of $\{\beta_1, \ldots, \beta_{p+q+1}\}$, where x consists of p disease determinants that are correctly measured in the validation study but will be mismeasured and denoted by X in the main study, and U_1 consists of q determinants that are always measured perfectly. In cohort studies of long duration and/or of relatively common outcomes and in cross-sectional studies of common symptoms, of which the current study is an example, the constant risk or prevalence ratio model is often of interest. This constant prevalence ratio (relative risk) model has the form

$$(3)\quad Pr(D=1|x,U_1) = f_1(D|x,U_1;\beta) = e^{\beta_0+\beta'_1x+\beta'_2U_1}.$$

If x were observed without error, as it is assumed to be in a validation study, the prevalence ratio would be estimated by multiplying $\hat{\beta}_1$ by an appropriate increment and exponentiating.

After selecting a suitable model for the relationship between the outcome of interest and exposure, the next step is to identify, in the validated subset of the study population $f_2(x|X,U_2;\theta)$, the measurement error/misclassification model that describes how the usual exposure measure, X, and other available variables, U_2, predict the gold standard, x. The covariates (U_1) in the relative risk model, f_1, may or may not be the same as the covariates (U_2) that determine the measurement error/misclassification process, f_2. Usually, there is no intrinsic interest in f_2; its form and parameters are the object of attention only to the extent that they need to be in order to obtain consistent, efficient estimates of β from the main study/validation study data. The measurement error/misclassification model should be fit to the data with all of the tools available to the adept analyst (i.e., diagnostics, graphics, goodness-of-fit tests, careful variable selection strategies, and so forth) in order to determine the best model of the measurement error process for the data at hand.

Using f_1 and f_2 (identified as described earlier), one next derives the model for the outcome, D, in the main study, denoted $f_3(D|X,U;\beta,\theta)$, as follows:

$$(4)\quad f_3(D|X,U;\beta,\theta) = \int f_1(D|x,U_1;\beta) f_2(x|X,U_2;\theta)\,dx,$$

where U contains the unique elements of U_1 and U_2. This derivation of f_3 is valid only when X is a surrogate for x, that is, when $f_1(D|x,X,U_1,U_2) = f_1(D|x,U_1)$. It can be shown that this assumption is equivalent to the assumption of nondifferential measurement error. In a cross-sectional study, measurement error may be differential. This possibility is less likely in a prospective study. The assumption of nondifferential measurement error can and should be empirically verified in a validation study.

All of the pieces are now in place to write down and maximize the likelihood function for the main study of size n_1 and the validation study of size n_2 to obtain $(\hat{\beta}_{ML}, \hat{\theta}_{ML})$:

$$(5)\quad L(\beta,\theta) = \sum_{i=1}^{n_1} \log[f_3(D_i|X_i,U_i;\beta,\theta)] + \sum_{i=n_1+1}^{n_1+n_2} \log[f_2(x_i|X_i,U_{2i};\theta)] + \sum_{i=n_1+1}^{n_1+n_2} \log[f_1(D_i|x_i,U_{1i};\beta)].$$

The first term gives the log likelihood for the main study, while the second and third terms give the log likelihood for the validation study. Standard likelihood theory provides several methods for estimating the variance of these estimates, calculating confidence intervals, and obtaining significance levels for hypothesis tests. Further technical details have been presented elsewhere.[19]

Results

In Table 1, we present the standard analysis from an uncorrected logistic regression model and results from the regression calibration measurement error correction procedure. The prevalence ratio increased after the regression calibration correction, with little increase in the confidence interval width.

Unlike the regression calibration method, the maximum-likelihood procedure made no a priori assumptions about the form of the measurement error model, $f_2(x|X,U_2;\theta)$. Rather, we used the data at hand to determine the model that best fit the data. Histograms of the exposure variables (x, X) are given in Figure 1. It could be seen that both exposure assessment technologies produced data that were sharply right-skewed with modes at 0. These features suggested a measurement error model that prohibits negative exposure values and allows for a high cumulative probability of x at low levels of X and suggested (but did not require) that the conditional distribution of x on (X,U_2) would be skewed in a manner similar to that for the marginal distributions.

A measurement error model assuming the normal distribution will produce a symmetric error distribution with equal probability of negative errors, a physical impossibility at low levels of x, where most of the data are concentrated. A log-normal measurement error model will produce a distribution with the desired long tail to the right of the mode, but this distribution has the unsuitable feature that, for values of x near and at 0 (where most of our data lie), there is no mass at 0. The gamma distribution has none of these disadvantages. If, at larger values of x, the distribution of the errors best fits a log-normal or normal distribution, the value of the shape parameter, ν, will mimic relevant aspects of these distributions (Figure 3). We thus fit our measurement error model, $f_2(x|X,U_2;\theta)$, to the family of gamma distributions and proceeded to find an empirically justified form for the parameters of this distribution, allowing for the possibility of either an additive or multiplicative relationship of these parameters as they depend on X.

We next investigated which covariates among those available needed to be included in $f_2(x|X,U_2;\theta)$, the measurement error model, using a liberal forward variable selection strategy. After inclusion of X, age, current smoking status, educational status, and type of work shift in the model, no other variables contributed to the fit.

Finally, using the specifications for f_1 and f_2 (as described earlier) and following equation 4, we derived

$$f_3(D=1|X,U) = e^{\beta_0+\beta'_2U_1}\left(\frac{\nu}{\nu-\mu\beta_1}\right)^{\nu},$$

where μ and ν were additive in X and the other covariates (U_2) (see Spiegelman and Casella[19] for further details of this deriva-

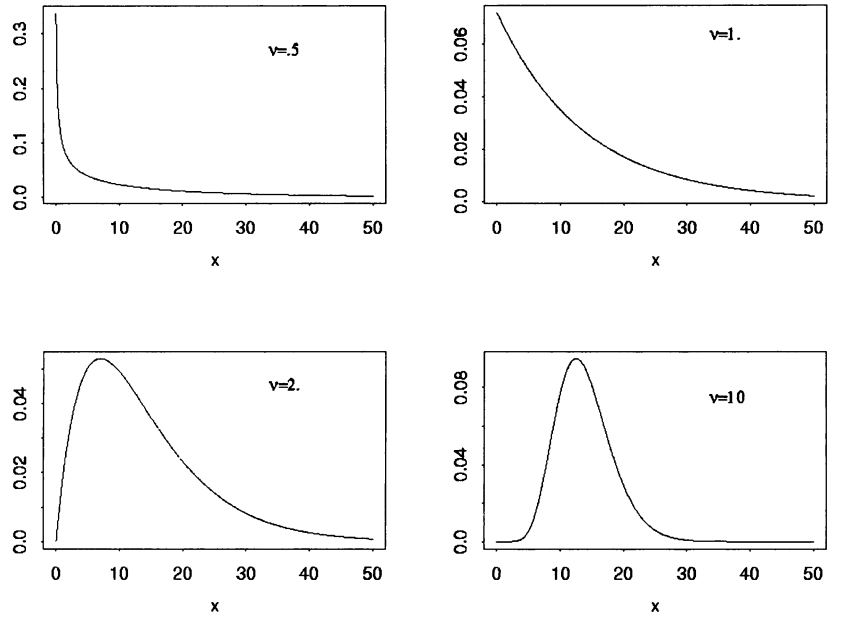

FIGURE 3—Shape of the gamma distribution (μ, ν) (μ = 13.875).

tion). We then used the likelihood function (equation 5) to select the components of U_1. Age, type of shift (SHIFT), and type of workplace (COMMHOSP) were each independent determinants of fever, given the measurement error model and the resulting form of f_3, where $f_1(D = 1|x, U_1) = e^{\beta_0 + \beta_1 x + \beta_{21} SHIFT + \beta_{22} COMMHOSP + \beta_{23} AGE}$.

The fourth row of Table 1 presents the results of the maximum-likelihood analysis with this gamma exposure measurement error model. A strong association between number of drugs mixed per week and fever was evident, with a 17% increase in the prevalence (95% CI = 4%, 26%) estimated for an increase over the interdecile range (difference between the 90th and 10th percentiles) of true exposure ($P = .03$). The estimate of the magnitude of this association was attenuated in the full data analysis using the constant prevalence ratio model (equation 3) and ignoring measurement error. When measurement error was ignored, the prevalence ratio was 6% (95% CI = 3%, 7%) for a 1-interdecile range increase in the number of drugs mixed per week.

Although uncertainty in the estimation of the parameters of the measurement error model and the uncertainty due to the presence of measurement error itself were fully accounted for, the likelihood-based correction for exposure measurement error led to no loss of statistical power for the hypothesis tested in these data; using the likelihood ratio test, the maximum-likelihood analysis yielded essentially the same P value as the uncorrected constant prevalence ratio model. In models with simple measurement error that follow the generalized linear model form both before and after measurement error is introduced, the P value from the uncorrected test is identical to the corrected one[32]; in general, however, there is no known result to suggest that the uncorrected test statistic will always have the same P value as the one obtained from the correctly specified likelihood.

A comparison of the results from the uncorrected logistic regression model and the uncorrected constant prevalence ratio model (see Table 1) indicated that the prevalence odds ratio was greater than the corresponding prevalence ratio. The prevalence odds ratio approximation to the prevalence ratio is likely to be an overestimate in this setting. In general, it is known that as the prevalence of the outcome of interest increases, the odds ratio approximation increasingly overestimates the prevalence ratio. The regression calibration procedure generated confidence intervals that were somewhat wider than those obtained from the maximum-likelihood analysis and a larger but nominally significant P value.

Discussion

The 2 methods presented in this paper make several critical assumptions. The first is that the outcome, along with all disease determinants for which validation data are unavailable, has been measured without error. If this assumption is false, validation data are needed for the variables in question. The 2 methods considered here require a second assumption: that a gold standard exists that can perfectly measure all model variables assumed to be measured with error and that these measurements were available in the validation study. It is known that the regression calibration method will provide unbiased point and interval estimates if the errors in the imperfect but unbiased ("alloyed") gold standard are uncorrelated with the errors in the usual method of exposure assessment.[33,34] When a perfect method of exposure assessment is unavailable, both of the measurement error correction procedures discussed in this paper can be interpreted as providing a method for estimating the point and interval estimates of relative risk that would have been obtained had the "alloyed" method been used to assess exposure in all study participants.

The second assumption is that measurement error is nondifferential (i.e., it does not depend on disease status). This is an assumption that can and should be verified empirically in a validation study. It is unlikely to be violated in prospective studies, in which exposure is measured at the start of follow-up; it is of greater concern in cross-sectional designs, many case–control studies, and some cohort studies, in which exposure histories are measured retrospectively or evolve during follow-up.

Finally, the 2 methods considered here assume that the measurement error process observed among the validation study participants is the same as that which occurred among the remaining study participants. If the validation study participants were randomly sampled from the main study, this assumption is satisfied. More generally, as long as sampling does not depend on the unobserved true exposure values conditional on (D, U′, X′)′, both the maximum-likelihood method and the regression calibration method will be valid.

In the Valanis et al. data, the maximum-likelihood analysis estimated an effect that was nearly 3-fold greater on the log prevalence ratio scale. The consequence of correction for measurement error on the point estimate in our example taken from the Valanis et al. study is less pronounced than would have been the case had the underlying prevalence ratio been of greater magnitude. An approximation roughly applicable to the constant prevalence ratio model and, approximately, to logistic

TABLE 1—A Comparison of Several Approaches to Estimation and Inference for the Valanis et al. Data on the Relationship between Fever in the Previous 3 Months and the Number of Antineoplastics Mixed per Week (n = 675)

Approach	$\hat{\beta}$	SE ($\hat{\beta}$)	Δ	Prevalence Ratio in Δ	95% Confidence Interval for Prevalence Ratio in Δ[a]	P[b]
Logistic regression, uncorrected (UCL)	0.00229	0.00089	IDR (X) = 52	1.13	1.03, 1.23	.01
Regression calibration correction to logistic regression (RC)	0.00580	0.00240	IDR (x) = 34	1.22	1.04, 1.43	.02
Constant prevalence ratio model, uncorrected (UC)	0.00165	0.00099	IDR (X) = 52	1.06	1.03, 1.07	.003
Constant prevalence ratio model, maximum likelihood (ML)	0.00464	0.00274	IDR (x) = 34	1.17	1.04, 1.26	.003

Note. ULC and RC are constant prevalence odds ratio models; UC and ML are constant prevalence ratio models. IDR = interdecile range.
[a]UCL and RC use Wald-type intervals; UC and ML use profile likelihood intervals.
[b]UCL and RC use Wald tests; UC and ML use likelihood ratio tests.

regression has been given as

$$RR = (RR_{UC})^{r^{-2}},$$

where RR is the true relative risk, RR_{UC} is the uncorrected relative risk, and r is the correlation coefficient between the gold standard and the usual measure of exposure assessment.[2] Here, $r = 0.70$; thus, for example, if the observed prevalence ratio were 2, the true prevalence ratio would have been approximately 4.11.

Although the uncorrected logistic regression analysis followed by the regression calibration correction procedure is easily undertaken by any public health investigator who has main study/validation study data, several requirements are embedded in this method that may or may not be acceptable for a given problem. A rare disease is needed at several theoretical junctures for unbiased estimation and inferences about risk ratios, and small measurement error and constant measurement error variance are also required.

In addition, the regression calibration procedure does not make full use of the data available; efficiency is lost relative to the analogous maximum-likelihood procedure. The maximum-likelihood procedure, in contrast, needs none of these requirements for unbiased and fully efficient estimation and inference. However, because the maximum-likelihood analysis is customized to the data at hand, it is difficult to provide standard software. For guidelines in terms of sample size and power calculations for cost-efficient main study/validation study designs, refer to Greenland[35] and Spiegelman and Gray.[36]

If the potentially substantial biases in estimation and inference due to exposure measurement error are truly to be accounted for, main study/validation studies should become the standard design paradigm, and measurement error methods such as those illustrated in this paper have much to offer when routinely incorporated in the analysis. □

Acknowledgments

This research was supported by National Institutes of Health grants OH00106, CA50597, and CA47727.

References

1. Willett W. *Nutritional Epidemiology*. New York, NY: Oxford University Press Inc; 1990.
2. Armstrong BK, White E, Saracci R. *Principles of Exposure Measurement in Epidemiology*. Oxford, England: Oxford University Press Inc; 1992.
3. Prentice RL, Sheppard L. Dietary fat and cancer: consistency of the epidemiologic data, and disease prevention that may follow from a practical reduction in fat consumption. *Cancer Causes Control*. 1990;1:81–97.
4. Savitz DA, Pearce NE, Poole C. Methodological issues in the epidemiology of electromagnetic fields and cancer. *Epidemiol Rev*. 1989;11:59–78.
5. *Health Risks of Radon and Other Internally Deposited Alpha-Emitters*. Washington, DC: National Academy Press; 1988.
6. Thomas D, Stram D, Dwyer J. Exposure measurement error: influence on exposure-disease relationships and methods of correction. *Annu Rev Public Health*. 1993;14:69–83.
7. Armstrong BG. The effects of measurement errors in relative risk regressions. *Am J Epidemiol*. 1990;132:1176–1184.
8. Clayton DG. Models for the analysis of cohort and case-control studies with inaccurately measured exposures. In: Dwyer JH, Lippert P, Feinleib M, Hoffmeister H, eds. *Statistical Models for Longitudinal Data Analysis*. Oxford, England; Oxford University Press Inc; 1991:301–331.
9. Chen TT. A review of methods for misclassified data in epidemiology. *Stat Med*. 1989;8:1095–1106.
10. Diamond EL, Lilienfeld AM. Effects of errors in classification and diagnosis in various types of epidemiological studies. *Am J Public Health*. 1962;52:1137–1144.
11. Bross I. Misclassification in 2×2 tables. *Biometrics*. 1954;10:478–486.
12. Rogot E. A note on measurement errors and detecting real differences. *J Am Stat* Assoc. 1961;56:314–319.
13. Willett WC, Hunter DJ, Stampfer MJ, et al. Dietary fat and fiber in relation to risk of breast cancer: an eight-year follow-up. *JAMA*. 1992;268:2037–2044.
14. Rimm EB, Ascherio A, Giovannuci EL, Spiegelman D, Stampfer MJ, Willett WC. Vegetable, fruit and cereal fiber intake and risk of coronary heart disease among men. *JAMA*. 1996;225:447–451.
15. Wacholder S, Dosemeci M, Lubin JH. Blind assignment of exposure does not always prevent differential misclassification. *Am J Epidemiol*. 1991;134:433–437.
16. Flegal KM, Keyl PM, Nieto FJ. Differential misclassification arising from nondifferential errors in exposure measurement. *Am J Epidemiol*. 1991;134:1233–1244.
17. Dosemeci M, Wacholder S, Lubin JH. Does non-differential misclassification always bias a true effect toward the null value? *Am J Epidemiol*. 1990;132:746–748.
18. Deddens JA, Hornung R. Quantitative examples of continuous exposure measurement errors that bias risk estimates away from the null. In: Smith CM, Christiani DC, Kelsey KT, eds. *Chemical Risk Assessment and Occupational Health: Current Applications, Limitations and Future Prospects*. Westport, Conn: Auburn House; 1994.
19. 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;53:395–409.
20. Valanis BG, Vollmer WM, Labuhn KT, Glass AG. Association of antineoplastic drug handling with acute adverse effects in pharmacy personnel. *Am J Hosp Pharm*. 1993;50:455–462.
21. Valanis BG, Vollmer WM, Labuhn KT, Glass AG. Acute symptoms associated with antineo-

plastic drug handling among nurses. *Cancer Nurs.* 1993;16:288–295.
22. Labuhn KT, Valanis BG, Glass AG, Vollmer WM. Implementing a cancer-control study in a National Cancer Institute clinical trials network. *Cancer Pract.* 1994;2:47–54.
23. Valanis B, Hertzberg V, Shortridge L. Antineoplastic drugs: handle with care. *Am Assoc Occup Health Nurs J.* 1987;35:487–492.
24. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates for non-random measurement error. *Stat Med.* 1989;8:1051–1069.
25. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error; the case of multiple covariates measured with error. *Am J Epidemiol.* 1990;132:734–745.
26. Armstrong B. Measurement error in the generalized linear model. *Commun Stat B Simula Computa.* 1985;14:529–544.
27. Carroll RJ, Spiegelman CH, Gordon KK, et al. On errors-in-variables for binary regression models. *Biometrika.* 1984;71:19–25.
28. Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika.* 1982;69:331–342.
29. Armstrong B, Whittemore AS, Howe GR. Analysis of case-control data with covariate measurement error: application to diet and colon cancer. *Stat Med.* 1989;8:1151–1163.
30. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. *Am J Epidemiol.* 1992;136;1400–1413.
31. Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. *Am J Epidemiol.* 1987;125:761–768.
32. Tosteson TD, Tsiatis AA. The asymptotic relative efficiency of score tests in a generalized linear model with surrogate covariates. *Biometrika.* 1988;75:507–514.
33. Wacholder S, Armstrong B, Hartge P. Validation studies using an alloyed gold standard. *Am J Epidemiol.* 1993;137:1251–1258.
34. Spiegelman D, Schneeweiss S, McDermott A. Measurement error correction for logistic regression models with an "alloyed gold standard." *Am J Epidemiol.* 1997;145:184–196.
35. Greenland S. Statistical uncertainty due to misclassification: implications for validation substudies. *J Clin Epidemiol.* 1988;41:1167–1174.
36. Spiegelman D, Gray R. Cost-efficient study designs for binary response data with Gaussian covariate measurement error. *Biometrics.* 1991;47:851–869.