Approaches to Uncertainty in Exposure Assessment in Environmental Epidemiology

Donna Spiegelman

Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115; email: stdls@channing.harvard.edu

Annu. Rev. Public Health 2010. 31:149–63

First published online as a Review in Advance on January 4, 2010

The *Annual Review of Public Health* is online at publhealth.annualreviews.org

This article's doi: 10.1146/annurev.publhealth.012809.103720

0163-7525/10/0421-0149$20.00

Key Words

measurement error, misclassification, validation study, regression calibration

Abstract

Uncertainty in assessment of individual exposure levels leads to bias, often, but not always, toward the null in estimates of health effects, and to underestimation of the variability of the estimates, leading to anticonservative p-values. In the absence of data on the uncertainty in individual exposure estimates, sensitivity analysis, also known as uncertainty analysis and bias analysis, is available. Hypothesized values of key parameters of the model relating the observed exposure to the true exposure are used to assess the resulting amount of bias in point and interval estimates. In general, the relative risk estimates can vary from zero to infinity as the hypothesized values of key parameters of the measurement error model vary. Thus, we recommend that exposure validation data be used to empirically adjust point and interval estimates of health effects for measurement error. The remainder of this review gives an overview of available methods for doing so. Just as we routinely adjust for confounding, we can and should routinely adjust for measurement error.

INTRODUCTION

Exposure: substance in the immediate environment of individuals, possibly associated with adverse health effects

Measurement error model: statistical model that specifies the relationship between true and surrogate individual continuous exposure measures

"We believe of the two of the major methodological issues raised in epidemiological studies of occupational exposures, that is, confounding and exposure misclassification, the latter is of far greater concern," concluded Blair et al. in their paper, *Methodological Issues Regarding Confounding and Exposure Misclassification in Epidemiological Studies of Occupational Exposures* (4). They continued, "It is rare to find substantial confounding in occupational studies (or in other epidemiologic studies for that matter), even by risk factors that are strongly related to the outcome of interest. On the other hand, exposure misclassification probably occurs in nearly every epidemiologic study." Reviews of the effects of exposure measurement error and the numerous methods that have been proposed to correct for biases that result when exposure measurement error is present have been published (1, 8, 10, 26, 46). There have been many theoretical investigations of the effects of measurement error on point and interval estimates of exposure-disease associations, and the widespread and profound presence of exposure measurement error in environmental health data has been documented (6, 12, 14, 32, 45). Nevertheless, there have been few original scientific publications that make use of existing methods for explicit exposure measurement error correction in environmental or occupational health fields, in contrast, for example, to nutritional epidemiology where such corrections are becoming more common (16, 31, 49).

The failure to apply existing methods for exposure measurement error correction in original environmental and occupational health research publications may be due to several factors. One reason is that the methods may be inappropriate for the particular features of environmental health studies. Existing methods for exposure measurement error correction typically assume normality of the exposure distribution, or additivity and homoscedasticity of the measurement errors, although some more general models have been considered, such as Lyles and Kupper's work on multiplicative measurement error models in the occupational epidemiology setting (28). Many strategies assume that the measurement error or misclassification process is fully known, which is never the case, or require a large internal validation study (see sidebar, Additional Definitions) is available, which is quite rarely the case. Recent work has addressed some of these deficits by explicitly providing methods for interval estimation (55), by considering methods that allow for heteroscedastic measurement error (D. Spiegelman, R. Logan, and D. Grove, submitted manuscript), and by directly modeling the joint empirical distribution of the exposure and its measurement error as it occurs in the data at hand (40). A second reason that the use of methods to correct for exposure measurement error and misclassification is not widespread is that the data required by these methods, the exposure validation data, may not be available. Frequently, however, exposure is validated, but the validation data are never

ADDITIONAL DEFINITIONS

Reliability study: In the infrequent instances when an average of replicate measures of exposure over a well-defined time interval approaches the individual's underlying true exposure, a reliability study collects one or more replicates of exposure from individuals internal or external to the main study

Sensitivity analysis: In the absence of validation data, or reliability data when the classical measurement error model holds, quantitative speculations are made about the extent of measurement error and used to assess impact on bias in health effect estimates; also called bias analysis and uncertainty analysis

Transportability: when the measurement error model in the external validation or reliability study can reasonably be assumed to be the same as the one that generated the exposure surrogate in the main study, the measurement error model parameters estimated from the validation or reliability substudy are transportable to the main study for valid measurement error correction

Validation study: a study in which data are simultaneously collected on the exposure surrogate and a gold standard method of exposure assessment. This study may be external to the main epidemiologic study, or be a subsample internal to the main study

explicitly used for correcting for measurement error bias. A third reason is related to the lack of human and technical capacity to perform the necessary adjustments. There is a lack of software to perform the calculations using existing methods, although software for some analysis methods (**http://www.hsph.harvard.edu/faculty/spiegelman/blinplus.html** and **http://www.hsph.harvard.edu/faculty/spiegelman/multsurr.html**) and study design methods (**http://www.mep.ki.se/%7Emarrei/software/**) is publicly available. In addition, there is a lack of human resources to implement these analyses among the current pool of biostatisticians and epidemiologists, who have not yet been adequately trained in the application of these methods.

In this review, we provide an overview of basic concepts required for understanding and applying statistical methods for the design and analysis of studies with mis-measured exposures. In addition, we illustrate some of the methods most relevant to occupational and environmental health research. In one example, we show that the estimated relative risk (95% CI) for the effect of environmental exposure to soot among Dutch schoolchildren goes from 2.2 (1.1–4.2) to 5.1 (1.0–25) after correction for bias due to exposure measurement error. In another example, we show that the precision of the estimated effect of NO_2 exposure on the prevalence of respiratory symptoms in schoolchildren improved by 34% through the application of appropriate exposure measurement error correction methods, which permitted use of all 2891 study participants instead of just the 1137 validation study subjects in the original publication. These examples demonstrate that proper application of measurement error correction methods reduces bias in effect estimates, usually away from the null, and increases precision in the estimates by using all of the data available, instead of either the main study alone or the validation study alone.

BASIC CONCEPTS

Exposure is the substance in the immediate environment of individuals and populations, but external to them (e.g., breathing zone concentration of a possible occupational hazard or air pollutant, concentration of radiation in food that is ingested). In environmental and occupational epidemiology, dose is defined as the concentration of the substance of interest in a target internal organ (2, 22). When the effect of exposure is of primary interest, and data on dose, rather than exposure, are more readily obtained, dose can be used as a surrogate marker for exposure that typically measures exposure with some error, unless it is a recovery biomarker (3), in which case it can be treated as the gold standard. A recovery biomarker is one that estimates the individual's true exposure level without bias, although it may include some random error. If the recovery biomarker contains random within-person (classical) measurement error, use of this variable in place of exposure will, in general, lead to bias in exposure effect estimates toward the null. If, instead, the effect of dose is of primary interest, then exposure is a surrogate for dose that typically measures dose with error. Often, the correlation between dose and exposure is low. It is sometimes said that exposure is the object of interest for public health research, whereas dose is the object of interest for investigating biological mechanisms. Henceforth, we assume, without loss of generality, that the health effects of exposure are the objects of interest.

Uncertainty analysis (5, 29), bias analysis (25), sensitivity analysis (13) (see sidebar), measurement error correction, and correction for misclassification are closely related enterprises. Measurement error correction refers to statistical methods that correct for bias in point and confidence interval (CI) estimates of exposure effects when the exposure is a continuous variable, whereas misclassification correction makes analogous bias corrections for categorical exposure variables. The statistical methods that can be used to correct for bias due to measurement error in continuous variables rarely can be used to correct for bias due to misclassification in categorical variables, and vice versa. Thus, the distinction between categorical and continuous exposures has

Dose: concentration of a substance of interest in a target internal organ

CI: confidence interval

Misclassification model: statistical model that specifies the relationship between true and surrogate individual categorical exposure measures

practical importance. These methods require a main study/validation study design, or when it is reasonable to assume the classical error model, i.e., that $Z_i = X_i + \varepsilon_i$ and $Var(\varepsilon_i) = \sigma_\varepsilon^2$, where X_i is the underlying true exposure for participant i and Z_i is the participant's observed (surrogate) exposure, a main study/reliability study (see sidebar) can be utilized. An example of an exposure in which a reliability study would be sufficient for measurement error correction is one in which the exposure of interest is individual average exposure to constituents of air pollution. In this case, daily air pollution constituent concentrations in the breathing zone of individual study participants are randomly sampled throughout the year, and as long as it is reasonable to assume that the underlying true exposure can be validly estimated by the average of the daily samples, replicate exposure data in a subsample are sufficient. **The statistical methods used must be determined by the measurement error model likely to characterize the relationship between the underlying true exposure, unobserved except in a validation study, and the surrogate, observed in the main study**. Which measurement error model that can reasonably be assumed to describe this relationship will dictate whether replicate data are sufficient or validation data are required and what statistical methods will provide valid point and interval estimates of effect following data collection. Methods are available for optimal choices of the size of both the main and validation studies in main study/validation study designs (15, 37, 41) and of the sizes of both the main and reliability studies, as well as the number of replicates in a main study/reliability study design (38).

When the validation or reliability study is internal to the main study, transportability (see sidebar) is guaranteed. When the measurement error in the validation or reliability study is assumed to be transportable to the main study, it is assumed that the same underlying measurement error model that generated the exposure data in the validation or reliability study generated the exposure data in the main study. The transportability assumption is not empirically verifiable when the validation or reliability study is external to the main study. If the features of the distribution (e.g., mean, variance, and shape of the histogram of the surrogate exposure in the main study) are similar to that in the validation or reliability study, and if other basic characteristics of the two study populations relevant to measurement quality are similar (e.g., age, sex, race, geographic region, and calendar time), the credibility of the transportability assumption is strengthened but not guaranteed. Many statistical methods for measurement error correction require estimates of features of the distribution of the true exposure given the surrogate exposure. Sometimes the full conditional distribution is required, but often only the conditional mean and variance are required. In general, statistical methods for measurement error correction that require features of the distribution of the surrogate exposure conditional on the true exposure, with or without features of the marginal distribution of the surrogate exposure in the main study, are most transportable.

When measurement error or misclassification correction is not possible, bias or sensitivity analysis can be employed to evaluate how the uncorrected results might change if measurement error or misclassification are taken into account. Bias analysis and sensitivity analysis are synonyms. In the point-wise version, the relative risk for measurement error or misclassification is corrected for bias by plugging into the formula for the true relative risk as a function of the observed relative risk estimate and the unobserved parameters that characterize the measurement error or misclassification model that likely generated the observed (surrogate) exposure, one set of hypothesized values of the key measurement error or misclassification model parameters at a time. The parameters to be plugged in are, in the case of binary misclassification, the sensitivity and specificity (21), or, in the case of classical measurement error, the intraclass correlation coefficient defined as the ratio of the between-subjects exposure variance over the total exposure variance. If it is not possible to derive a closed-form

expression for the true relative risk as a function of the observed relative risk estimate and the unobserved parameters that characterize the measurement error or misclassification model that likely generated the observed (surrogate) exposure, it is sometimes possible to derive the likelihood function for the available data as a function of the assumed relative risk model and assumed measurement error or misclassification model. Then, a realistic range of values for the unknown parameters of the measurement error or misclassification model can be plugged into this expression, one by one, and the range of relative risks that would be obtained if the measurement error or misclassification model were fully known and were characterized by the hypothesized values can be calculated. A more sophisticated approach, Bayesian (13) or probabilistic (25) sensitivity analysis, generalizes the point-wise sensitivity analysis previously described by using a weighted distribution of possible measurement error or misclassification parameter values in place of a single point. Then, the median most likely assumed value for the relative risk can be obtained, along with, for example, the interquartile range.

If validation or reliability data are not available, the parameter of interest, typically the relative risk for the outcome in relation to a standard incremental difference in exposure levels, cannot be validly estimated from the data at hand but, rather, is estimated with bias, often enough bias to change the fundamental interpretation of the data (56). With the classical measurement error model and binary misclassification, it is well known that the estimated relative risk is biased toward the null (13). That is, as long as measurement error or misclassification is nondifferential, the relative risk relating true exposure to the outcome is bounded by the estimated value and one. Many epidemiologic papers do not go further, stating that if the misclassification is nondifferential, bias is toward the null. The parameters of a nondifferential measurement error model do not vary with outcome. Recall bias in a case-control study is a classic example of differential misclassification. If measurement error or misclassification is differential, bias in the relative risk can be in either direction.

With binary misclassification or classical measurement error, hypothesis tests are valid without measurement error correction (24, 48). Hence, if hypothesis testing is the sole purpose of a study, no validation or reliability data are needed. This will seldom be the case in environmental and occupational epidemiology; rather, point and interval estimation of exposure effects are the primary aim of most studies. Even under the simplest measurement error and misclassification models, bias in point and interval estimates will occur. Bias in point and interval estimates increases exponentially as measurement error or misclassification worsens (56). Occupational and environmental exposure data are measured with error, often substantial error (6, 12, 14, 56). Thus, there will rarely be instances in occupational and environmental research when measurement error correction is not needed to obtain valid point and interval estimates.

We present several examples of the successful application of methods to correct for exposure measurement error in main study/validation study designs.

Regression calibration: method used to correct point and interval estimates of the health effects of exposure for measurement error, using data from a suitable validation study

Example 1: Regression Calibration in a Main Study/External Validation Study Design—The Dutch Study of the Respiratory Effects of Outdoor Air Pollution Constituents

In this study (52), we applied the regression calibration method to adjust for exposure measurement error in outdoor air pollutant exposure in relation to chronic respiratory symptoms and other health outcomes among schoolchildren living near freeways (18, 19, 51, 53). We calculated the relative risk of wheeze, conjunctivitis, phlegm, and elevated total serum immunoglobulin E in relation to outdoor air pollution concentrations measured at the school and adjusted for exposure measurement error using the regression calibration method (17, 34, 35). This method requires a main study in which data on the health outcome, surrogate exposure,

OR: odds ratio

and confounding covariates are available and a validation study in which true exposure, surrogate exposure, and confounding covariates are assessed. We considered the outdoor concentrations of the air pollutants to be the surrogate exposures and the personal exposures the true exposures. We focused on the health effects of exposure to soot. The main study included 2083 children from 24 schools located within 400 meters of freeways in the Netherlands (18). Measurements of soot were conducted at all 24 schools between April, 1997 and July, 1998 (19). The validation study, external in design, was conducted six years after the main study, when personal soot measurements were collected in 54 Dutch schools (53). Outdoor measurements at the schools were performed concurrently with personal measurements.

Rosner et al.'s regression calibration method for adjustment of point and interval estimates for bias due to exposure measurement error is a three-step procedure (34). First, the unadjusted point estimates and their variances are obtained by fitting the standard regression model in the main study, in this case, a log-binomial model. Log-binomial models were employed to directly estimate the prevalence ratio, rather than the prevalence odds ratio (OR), because the prevalence OR is an approximation to the prevalence ratio that fails when the disease is not rare (42, 54). This example considers some such outcomes. The log-binomial model that is of interest in this analysis has the form $\Pr(D=1|E, C_1, \ldots, C_s) = e^{\beta_0+\beta_1 E+\beta_{21}C_1+\cdots+C_{2s}}$, where, for example, when the outcome of interest is conjunctivitis and the exposure of interest is soot, $D = 1$ if the child reported conjunctivitis and 0 otherwise; E = personal exposure to soot ($\mu g/m^3$); and $C_1, \ldots, C_s$ are the $s = 6$ confounders, parental smoking, gas cooking, presence of an unvented water heater, sex, age (years), and current pet possession. The relative risk is thus $\frac{\Pr(D=1|E+\Delta, C_1, \ldots, C_s)}{\Pr(D=1|E, C_1, \ldots, C_s)} = e^{\beta_1 \Delta}$, where Δ is a scientifically meaningful increment in the exposure of interest. Thus, in this case, the first step is to fit a log-binomial regression of the outcome of interest, here, wheeze, on the surrogate exposure for soot and the other covariates, assumed measured without error. Second, in the validation study, measurement error model parameters are estimated by regressing the true exposure on the surrogate exposure and all other covariates included in the primary regression model, following the standard linear model $E = \gamma_0 + \gamma_1 X + \gamma_{21} C_1 + \cdots \gamma_{2s} C_s + \varepsilon$, where X is the validation study subject's surrogate exposure value and ε is a random error term. Third, estimates are adjusted for measurement error by combining these two sets of estimates and their variance-covariances to obtain point and interval estimates of relative risk corrected for measurement error. The regression calibration method has been extended to apply to linear regression (43), survival data analysis (43, 57), and relative risk regression (17).

All statistical methods require certain conditions to be met for the results to be valid. For the regression calibration method to be valid, we need to verify that the assumptions associated with it are applicable to the data at hand. These include the following (34): (Assumption 1) The measurement error model is linear and homoscedastic; (Assumption 2) the main study model is linear on the assumed scale, in this case, on the log scale; (Assumption 3) measurement error is not severe, or the disease is rare and the error term in the measurement error model is normally distributed; (Assumption 4) the surrogate exposure contains no further information about the distribution of disease once data on the true exposure is available; (Assumption 5) if the exposure validation method contains error, these errors are uncorrelated with the errors in the surrogate exposure variable; and (Assumption 6) transportability, meaning the measurement error model in the validation study can be reasonably assumed to be the model that generated the main study surrogate exposure data, at least with respect to the critical parameters used in the procedure. Importantly, if measurement error is not severe, normality is not required for valid application of this method. Most of the assumptions were empirically verifiable in this study using standard statistical diagnostic procedures.

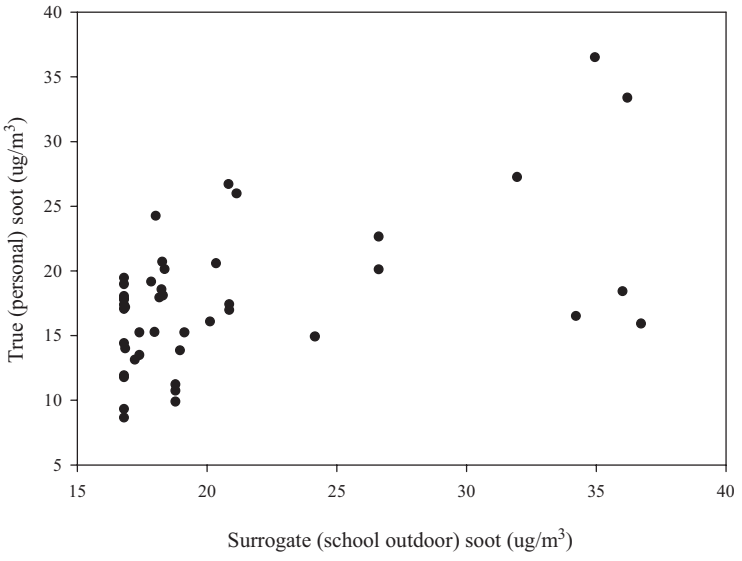

Figure 1

Scatter plot of average personal soot exposure versus outdoor school in the Dutch validation study ($n = 45$); $r = 0.53$.

Although the validation study was conducted six years after the main study and in different Dutch schools, the same exposure assessment method was evaluated in similar environments as those in the main study, and the investigators believe that transportability can be reasonably assumed. Further details regarding the verification of the assumptions have been given (53).

Figure 1 shows the scatterplot of the personal level of soot versus the school's outdoor concentration in the validation study. The correlation coefficient between the true and surrogate exposure was 0.53. The measurement error model is shown in **Table 1**. The unadjusted estimate of the prevalence ratio for conjuctivitis was 2.18 (95% CI, 1.13–4.21), and after adjusting for measurement error, the estimate increased to 5.06 (95% CI, 1.02–24.96). This suggests that the effect of soot exposure on risk of conjunctivitis was underestimated by more than half in the original analysis. Effects of measurement error adjustment for current conjunctivitis, current phlegm, and elevated total IgE were similar. The larger CIs reflect the small sample size of the validation study, as well as the increased estimate of uncertainty due to the magnitude of the measurement error observed. Software

Table 1 Measurement error model for average personal exposure to soot ($\mu g/m^3$) ($n = 45$)

Variable name	$\hat{\gamma}$	$S\hat{E}(\hat{\gamma})$	*p*-value
Intercept	13.16	12.48	0.30
School outdoor soot ($\mu g/m^3$)	0.48	0.13	<0.001
Exposure to ETS (yes/no)	5.72	1.80	<0.01
Exposure to gas cooking (yes/no)	2.55	1.76	0.16
Presence of unvented water heater in the kitchen (yes/no)	−1.84	4.81	0.70
Sex (boy/girl)	0.19	1.42	0.89
Age (year)	−0.80	1.01	0.43
Current pet possession (yes/no)	1.62	1.37	0.25

$\hat{\gamma}$ = regression slope; $S\hat{E}(\hat{\gamma})$ = standard error. For all categorical variables, yes is coded as 1 and no as 0.

Exposure surrogate: usual measure of exposure in an environmental epidemiology study, typically assessed with error, relative to the underlying true value

implementing this method is available on the author's Web sites (**http://www.hsph.harvard.edu/faculty/spiegelman/blinplus.html** and **http://www.hsph.harvard.edu/faculty/spiegelman/relibpls8.html**).

Example 2: Regression Calibration with Multiple Surrogates—The Six Cities Study of the Respiratory Effects of Indoor Exposure to NO_2

In this example, the effect of indoor NO_2 exposure on the annual prevalence of lower respiratory symptoms was evaluated by applying a regression calibration measurement error correction method that efficiently incorporates information on multiple exposure surrogates for an exposure of interest, as is common in environmental health research (55). This method has several advantages compared to the standard analysis uncorrected for measurement error. All of the available data are used, bias due to measurement error is removed, the relative risk is scaled to the measured units of the exposure of interest, and the variance of the estimate of the exposure effect is adjusted for the additional uncertainty inherent in the study, which is due to the estimation of exposure among those also without NO_2 levels. A SAS macro for this procedure is available at **http://www.hsph.harvard.edu/faculty/spiegelman/multsurr.html**. The effect of NO_2 exposure on lower respiratory symptoms was assumed to follow the logistic regression model,

$$(\Pr[D_i = 1]) = \beta_0 + \beta_1 E_i + \beta_{21} C_{1i} + \cdots + \beta_{2s} C_{si}, \qquad 1.$$

where for each subject i, D_i, is a binary indicator variable that takes on value 1 if lower respiratory symptoms occurred during the past year, and 0 otherwise; E_i is the quantitative measure of NO_2 level; and $C_{1i}, \ldots, C_{si}$ are the values of covariates assumed to be perfectly measured (30). The parameter β_1 is the log OR for a one-unit increment in the quantitative exposure and is the parameter of interest in this analysis. To apply this method, a logistic regression model was fit in the main study using data on all r surrogates, $X_1, \ldots, X_r$, and the s perfectly measured covariates, $C_1, \ldots, C_s$, $\log it(\Pr[D_i = 1]) = \alpha_0 + \alpha_{11} X_{1i} + \cdots + \alpha_{1i} X_{ri} + \alpha_{21} C_{1i} + \cdots + \alpha_{2s} C_{si}$, for subject i, $i = 1, \ldots, n_1$. Then, the quantitative exposure measure was modeled as a linear function of the exposure surrogates and the covariates in the validation study:

$$E_i = \gamma_0 + \gamma_{11} X_{1i} + \cdots + \gamma_{21} X_{ri} + \gamma_{21} C_{1i} + \cdots + \gamma_{2s} C_{si} + \varepsilon_i, \qquad 2.$$

where ε_i are independent random errors with mean 0 and constant variance σ^2 and $i = 1, \ldots, n_2$. A bias-corrected estimate is obtained by computing the log relative risk estimates for each of the r surrogate exposure variables following $\hat{\beta}_{1j} = \hat{\alpha}_{1j}/\hat{\gamma}_{1j}$, $j = 1, \ldots, r$, then combining them using inverse variance weights (39). This results in a single estimate, $\hat{\beta}_1$, which has minimum variance and is in the original units of the quantitative exposure. The variance of $\hat{\beta}_1$ was derived using the multivariate delta method. When $r = 1$, the measurement error-corrected point estimate of the log of the relative risk reduces to the original regression calibration estimate. When β_1 can be estimated in an (internal) validation study, as in this case, this estimate can be efficiently combined with the estimate described above, $\hat{\beta}_1$, to produce a single summary estimate (39).

This study included $n_1 = 1754$ main study subjects and $n_2 = 1137$ validation study subjects. Seven surrogates for the NO_2 exposure were used: NO_2 sources included the presence of a gas stove with or without a pilot light, the presence of a kerosene space heater, the presence of a wood stove, and the usage of stove for heating; residential characteristics included fan usage for kitchen ventilation and the total number of rooms in the home. The confounders included parental characteristics, such as history of respiratory diseases, employment, smoking information, education, and marital status. The measurement error model (Equation 2) was fit to the validation data, from which was obtained the deattenuation factors and weights for each surrogate. Results

Table 2 Six Cities study: NO_2 exposure in relation to annual prevalence of respiratory symptoms[a]

Variable	Uncorrected analysis ($n_1 = 1754$)	Measurement error corrected ($n_1 = 1754$)		Validation study alone ($n_2 = 1137$)	Combined analysis ($n_1 + n_2 = 2891$)
	$\widehat{OR}$ [95% *CI*][b]	$\widehat{OR}$ [95% *CI*]	Weight	$\widehat{OR}$ [95% *CI*]	$\widehat{OR}$ [95% *CI*]
NO_2 (per 15 ppb increment)				1.41 [1.13, 1.75]	1.45 [1.20, 1.75]
Surrogates (**W**)		1.60 [1.10, 2.32]			
Gas stove, no pilot	0.68 [0.42, 1.10]	0.94 [0.86, 1.02][d]	0.18		
Gas stove, pilot	1.54 [0.94, 2.52]	1.04 [0.99, 1.10]	0.51		
Stove heater	1.61 [1.05, 2.47]	1.50 [0.72, 3.14]	0.002		
Fan	0.93 [0.81, 1.07]	1.11 [0.90, 1.36]	0.03		
Wood stove	0.91 [0.66, 1.25]	1.05 [0.89, 1.27]	0.04		
Number of rooms in the home	0.99 [0.92, 1.06][c]	1.04 [0.85, 1.27]	0.03		
Kerosene heater	1.41 [0.96, 2.07]	1.07 [0.99, 1.16]	0.20		

[a] Adjusted for cities, single marital status, higher education status, parental history of bronchitis or emphysema, parental history of asthma, gender, age, and the total packs of cigarette smoked inside the child's home.
[b] $\widehat{OR}$s and their 95% CIs are given for the effect of each surrogate, adjusted for all others and for the model covariates.
[c] $\widehat{OR}$s and their 95% CIs are given for the effect of a one-room increase, adjusted for all other surrogates and for the model covariates.
[d] $\widehat{OR}$s and their 95% CIs are given per 1 ppb increment in NO_2.

are given in **Table 2**. The weights describe the extent to which each surrogate contributes to the final summary estimate. Using these weights, the measurement error-corrected estimated exposure effect combining the individual surrogates gave an $\widehat{OR} = 1.60$ (95% CI = [1.10,2.32]) for a 15 ppb increment in NO_2. This was consistent with results obtained by fitting the model (Equation 1) among the validation subjects alone ($\widehat{OR} = 1.41$, 95% CI = [1.13,1.75]). Combining the results from the modified regression calibration approach and from the logistic regression model in the validation subjects alone, a 15 ppb increment in NO_2 was associated with a 50% increase in the odds of having one or more lower respiratory symptoms (95% CI 1.20, 1.75). The precision of $\widehat{OR}$ increased by 34% in the combined analysis using all available data. The SAS code that produced this analysis has been published (27).

Example 3: Regression Calibration with Heteroscedastic Measurement Error Model Variance—The ACE Study of Occupational Exposure to Chemotherapeutic Agents

Many environmental and occupational exposures, such as indoor NO_2 exposure (**Figure 2**), have a skewed distribution. This common feature often, although by no means always, leads to a heteroscedastic variance in the measurement error model (Equation 2), violating the first assumption for valid application of the regression calibration method. In this case, we extend the regression calibration method to allow for heteroscedastic measurement error. Let $\Pr(D = 1|E) = \frac{e^{(\beta_0+\beta_1 E)D}}{1+e^{\beta_0+\beta_1 E}} \approx e^{(\beta_0+\beta_1 E)D}$ under the rare disease assumption. Let $f_2(E|X)$, a general expression for the measurement error model, follow a normal distribution with conditional mean $\alpha' + \gamma X$ as given by Equation 2, and conditional variance $Var(E|X) = h(X)\sigma^2$. Then, by completing squares,

$$\begin{aligned}\Pr(D = 1|X) &= \int_E \Pr(D = 1|E) f_2(E|X) dE \\ &\approx \int_E \frac{e^{\beta_0+\beta_1 E - \frac{1}{2h(X)\sigma^2}(E-\alpha'-\gamma X)^2}}{\sqrt{2\pi h(X)\sigma^2}} dE \\ &= \cdots = e^{\beta_0+\beta_1\alpha'+\beta_1\gamma X+\sigma^2\beta_1^2 h(X)} \\ &= e^{\beta_0^*+\beta_{11}^* X+\beta_{12}^* h(X)}. \qquad 3.\end{aligned}$$

Kuha derived the distribution of $D|X, C_1, \ldots, C_s$ for logistic regression with rare disease and multivariate normality for the heteroscedastic distribution of the exposure

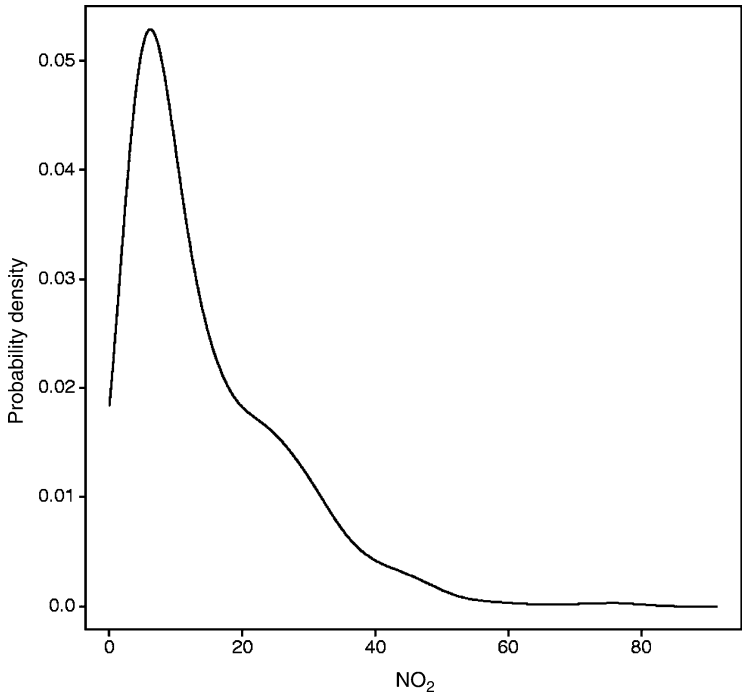

Figure 2

Empirical distribution of NO_2 ($n_2 = 1137$).

(23). A similar result for general mean-variance models, using a second-order Taylor series expansion around the conditional mean of E given X, has been given previously (8, 9), and for survival data analysis (33) and for linear regression (11), where Equation 3 is exact.

The result in Equation 3 leads to the following procedure for obtaining $\hat{\beta}_{RCH}$: *a*) a logistic regression model of D on $(X, C_1, \ldots, C_s)$ and $b(X)$ is run in the main study to obtain $\hat{\beta}^*_{11}$ and $\hat{\beta}^*_{12}$ and their estimated variances; *b*) a weighted linear regression is run in the validation study, with weights $1/b(X)$, to obtain $\hat{\sigma}^2$ and $\hat{\gamma}$; and *c*) $\hat{\beta}_{11}$ and $\hat{\beta}_{12}$ are obtained by solving for $\hat{\beta}^*_{11}$ and $\hat{\beta}^*_{12}$ for β using $\hat{\sigma}^2$ and $\hat{\gamma}$ as indicated by Equation 3 above and efficiently combined to produce a single estimate, $\hat{\beta}_{RCH}$, $\hat{\beta}_{RCH} = w_1\hat{\beta}_{11} + (1 - w_1)\hat{\beta}_{12}$. The asymptotically minimum variance weights and their derivation, as well as the formula for the variance of $\hat{\beta}_{RCH}$, are given in Appendix 1.

This method was applied to a cross-sectional study of acute health effects from occupational chemotherapeutics exposure in 675 hospital pharmacists (50). Average weekly chemotherapeutics exposure (X) was self-reported on a questionnaire; in a subsample of 56 pharmacists, on-site drug mixing diaries were kept for one to two weeks (E). The correlation between these two methods of exposure assessment was 0.70. The research objective was estimation and inference about the prevalence OR for acute health effects related to chemotherapeutics exposure. In this case, we focused on fever prevalence in relation to exposure. There were 110 cases of fever. The correlation between the absolute value of the residuals and the predicted values from the linear measurement error model (7) for diary data (E), conditional upon the questionnaire data (X) and other model covariates, was 0.21, indicative of moderate heteroscedasticity. Two outliers were evident. When these were removed, the

correlation between the absolute value of the measurement error model regression residuals and the predicteds was 0.26. These analyses and the ones that follow all adjusted for three covariates: age in years, work shift (1 if night or rotating shift, 0 if day shift), and employed by a community hospital (1 if yes, 0 if no).

To apply this extension to regression calibration for heteroscedasticity, we first needed to identify a form for $h(X)$. We searched through the class of functions $h(X) = (X + b)^p$ to find the transformation in this class for which the correlation between the absolute value of the weighted residuals from the weighted least squares regression of E on X and the other covariates is nearest to zero, where the weights are $h^{-1}(X)$. For the full validation data set, including apparent outliers, the best form for $h(X)$ was thus identified as $h(X) = X$. Excluding the two outliers, we found that the optimal transformation was estimated as $h(X) = X^{1.5}$, but this choice of $h(X)$ gave an unrealistic estimate for the OR of 0.41 (95% CI 0.33, 0.50). Therefore, we chose the nearly optimal transformation, $h(X) = X$, as was suggested by analysis in the full validation study, and obtained a measurement-error corrected value of 1.24 (95% CI 1.05–1.48) for the extreme quintile contrast in daily drug exposure, which was very similar to the value obtained from the standard regression calibration method that assumes measurement error model homoscedasticity. A simulation study confirmed these findings—in all cases considered in the simulation study, the standard regression calibration estimator had less empirical bias and better empirical CI coverage than the heteroscedasticity-adjusted version, even when the measurement error model heteroscedasticity was severe. The additional complexity of the heteroscedasticity-adjusted version led to more finite sample bias than was removed by the extra calculations. Although regression calibration formally requires a homoscedastic measurement error variance, in many realistic applications, little bias is induced when this assumption is moderately violated. Again here, measurement error correction led to substantial deattenuation because the uncorrected $\widehat{OR}$ (95% CI) was 1.08 (1.02–1.15).

CONCLUSIONS

Bias due to exposure measurement error is a major limitation in the validity of occupational and environmental studies. Methods have been developed that accommodate the features of study design and data distributions found in these studies. These methods implement explicit adjustments for this source of bias, using the exposure validation study to characterize the magnitude and other features of the measurement error. Both point and interval estimates of effect are adjusted. Methods address several commonly occurring situations in environmental and occupational epidemiology, including multiple surrogates for a single mis-measured exposure and heteroscedastic measurement error. User-friendly SAS macros are available to implement many of these methods. Papers have been published applying these methods to the analysis of occupational and environmental studies (17, 20, 27, 44, 47, 52, 55). Empirical adjustment for measurement error and misclassification in environmental and occupational health using main study/validation study designs is preferable over uncertainty analysis, bias analysis, and sensitivity analysis whenever internal or external validation data are available. Just as we routinely adjust for confounding, we can and should routinely adjust for measurement error.

SUMMARY POINTS

1. Bias due to exposure measurement error is a major limitation to the validity of environmental health studies. It is likely *the* major limitation.

2. When suitable validation data are available, methods are available to empirically adjust point and interval estimates of health effects for bias due to exposure measurement error.
3. Empirical adjustment for measurement error and misclassification in environmental and occupational health using main study/validation study designs is preferable over uncertainty analysis, bias analysis, and sensitivity analysis whenever internal or external validation data are available.
4. To obtain valid results, the measurement error or misclassification model assumed by a statistical method to adjust for measurement error or misclassification must be reasonable. Empirical verification of assumptions is best, to the greatest extent possible.
5. Just as we routinely adjust for confounding, we can and should routinely adjust for measurement error. To do so, environmental health researchers should routinely validate their exposure measures, ideally in a subsample of the main study population.

FUTURE ISSUES

1. If use of the bias correction methods described in this article are to become standard, more comprehensive and widespread training of biostatisticians, epidemiologists, and environmental health scientists is needed in these methodologies.
2. Further development of measurement error bias correction methods that apply to survival data analysis of prospective cohort studies with time-varying covariates such as cumulative and distributed lag exposure variables is needed.
3. Further development of measurement error bias correction methods that are valid for models with nonlinear exposure-response relationships, thresholds, and latency periods is needed. Little is presently known about the effects of exposure measurement error on assessment of nonlinearity, or estimates of thresholds or latency periods.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

APPENDIX 1: DERIVATION OF THE OPTIMAL WEIGHTS AND THE VARIANCE FOR $\hat{\beta}_{RCH}$

Let $V_1 = \text{Var}(\hat{\beta}_{11})$, $V_2 = \text{Var}(\hat{\beta}_{12})$, and $V_{12} = \text{Cov}(\hat{\beta}_{11}, \hat{\beta}_{12})$. By the multivariate delta method, $V_1 = \frac{Var(\hat{\beta}^*_{11})}{\gamma^2} + \frac{(\beta^*_{11})^2}{\gamma^4} Var(\hat{\gamma})$. Under the heteroscedastic measurement error model, when γ is estimated using weighted linear regression with weights $h(X_i)$, $i = 1, \ldots, n_2$, $Var(\hat{\gamma}) \approx \sigma^2[\boldsymbol{X}' diag\{\frac{1}{\text{h}(X_i)}\}\boldsymbol{X}]^{-1}$, where $\boldsymbol{X}$ is an $n_2 \times 1$ vector with elements X_i, $i = 1, \ldots, n_2$ (36). Again, by the multivariate delta method, $V_2 \approx \frac{Var(\hat{\beta}^*_{12})}{2\sigma^2 |\beta^*_{12}|} + \frac{|\beta^*_{12}| Var(\hat{\sigma}^2)}{2\sigma^6}$, because by arguments analogous to those given in Appendix 1 of Spiegelman et al. (39), $Cov(\hat{\beta}^*_{12}, \hat{\sigma}^2)$ is asymptotically 0, and $V_{12} \approx sign(\beta^*_{11})\, sign(\beta^*_{12}) \frac{Cov(\hat{\beta}^*_{11}, \hat{\beta}^*_{12})}{\gamma\sigma\sqrt{2|\beta^*_{12}|}}$, where $sign(t) = 1$ if t is positive and -1 if t is negative. All

other covariance terms are 0, because by the Gauss-Markov theorem, $Cov(\hat{\gamma}, \hat{\sigma}^2) = 0$ 1, and by arguments analogous to those given in Appendix 1 of Spiegelman et al. (39),

$$Cov\left[\begin{pmatrix} \hat{\beta}^*_{11} \\ \hat{\beta}^*_{12} \end{pmatrix}, \begin{pmatrix} \hat{\gamma} \\ \hat{\sigma}^2 \end{pmatrix}\right] \approx \mathbf{0}.$$

Thus, $Var[w_1 \hat{\beta}_{11} + (1 - w_1) \hat{\beta}_{12}] = w_1^2 V_1 + (1 - w_1)^2 V_2 + 2 w_1(1 - w_1) V_{12}$.

The estimates of V_1, V_2, and V_{12} are obtained by substituting the parameters for their estimates for the uncorrected primary regression of Y on $[X,b(X)]$ in the main study, and the weighted linear regression of x on X in the validation study. Likewise, the variances of these parameter estimates are obtained from these same regression analyses, and substituted into the expressions for V_1,V_2, and V_{12} to obtain estimates of these quantities.

To derive the optimal weight, w_1, we need to minimize

$$g(w_1) = w_1^2 V_1 + (1 - w_1)^2 V_2 + 2w_1(1 - w_1)V_{12}$$

with respect to w_1, because $w_2 = 1 - w_1$ to obtain a consistent estimator. The single extremum of this function, subject to the constraint, is at $w_1 = \frac{V_2 - V_{12}}{V_1 + V_2 - 2V_{12}}$. This is a global minimum as long as $V_1 + V_2 > 2V_{12}$, a condition that is likely to be met in most situations. With these optimal weights, the variance of $\hat{\beta}_{RCH}$ is thus

$$Var(\hat{\beta}_{RCH}) = \frac{V_1 V_2 - V_{12}^2}{V_1 + V_2 - 2V_{12}}.$$

To estimate $Var(\hat{\beta}_{RCH})$, estimates of V_1, V_2, and V_{12} are obtained from the fit of Equation 1 (above) to the main study data, $Var(\hat{\gamma})$ is estimated by plugging $\hat{\sigma}^2$ into the expression for $Var(\hat{\gamma})$ given above, and $Var(\hat{\sigma}^2) = 2\sigma^4/[n_2 - dim(\gamma) - 1]$.

LITERATURE CITED

1. Armstrong BG. 1990. The effects of measurement errors on relative risk regressions. *Am. J. Epidemiol.* 132:1176–84
2. Axelson O, Westberg H. 1992. Introductory note on the concepts of exposure and dose in occupational epidemiology. *Am. J. Ind. Med.* 21:3–4
3. Bingham SA. 2002. Biomarkers in nutritional epidemiology. *Public Health Nutr.* 5:821–27
4. Blair A, Stewart P, Lubin JH, Forastiere F. 2007. Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures. *Am. J. Ind. Med.* 50:199–207
5. Bogen KT. 1990. *Uncertainty in Environmental Health Risk Assessment (Environment – Problems and Solutions)*. New York: Garland Publ., Inc. 195 pp.
6. Brunekreef B, Noy D, Clausing P. 1987. Variability of exposure measurements in environmental epidemiology. *Am. J. Epidemiol.* 125:892–98
7. Carroll RJ, Ruppert D. 1988. *Transformation and Weighting in Regression (Monographs on Statistics and Applied Probability)*. London: Chapman & Hall. 264 pp.
8. **Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. 2006. *Measurement Error in Nonlinear Models: A Modern Perpective*. London: Chapman & Hall. 488 pp. 2nd ed.**
9. Carroll RJ, Stefanski LA. 1990. Approximate quasi-likelihood estimation in models with surrogate predictors. *J. Am. Stat. Assoc.* 85:652–63
10. Chen TT. 1989. A review of methods for misclassified categorical data in epidemiology. *Stat. Med.* 8:1095–106
11. Fuller WA. 1987. *Measurement Error Models*. New York: Wiley. 440 pp.
12. Gladen B, Rogan WJ. 1979. Misclassification and the design of environmental studies. *Am. J. Epidemiol.* 109:607–16

8. Comprehensive technical treatment of methods for adjusting for measurement error in nonlinear regression models.

13. Gustafson P. 2003. *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments*. Boca Raton, FL: Chapman & Hall/CRC. 200 pp.

14. Hatch M, Thomas D. 1993. Measurement issues in environmental epidemiology. *Environ. Health Perspect.* 101(Suppl. 4):49–57
15. Holcroft CA, Spiegelman D. 1999. Design of validation studies for estimating the odds ratio of exposure-disease relationships when exposure is misclassified. *Biometrics* 55:1193–201
16. Holmes MD, Stampfer MJ, Wolf AM, Jones CP, Spiegelman D, et al. 1998. Can behavioral risk factors explain the difference in body mass index between African-American and European-American women? *Ethn. Dis.* 8:331–39
17. Horick N, Weller E, Milton DK, Gold DR, Li R, Spiegelman D. 2006. Home endotoxin exposure and wheeze in infants: correction for bias due to exposure measurement error. *Environ. Health Perspect.* 114:135–40
18. Janssen NA, Brunekreef B, van Vliet P, Aarts F, Meliefste K, et al. 2003. The relationship between air pollution from heavy traffic and allergic sensitization, bronchial hyperresponsiveness, and respiratory symptoms in Dutch schoolchildren. *Environ. Health Perspect.* 111:1512–18
19. Janssen NAH, van Vliet PHN, Aarts F, Harssema H, Brunekreef B. 2001. Assessment of exposure to traffic related air pollution of children attending schools near motorways. *Atmos. Environ.* 35:3875–84
20. Keshaviah AP, Weller E, Spiegelman D. 2003. Occupational exposure to methyl tertiary butyl ether in relation to key health symptom prevalence: the effect of measurement error correction. *Environmetrics* 14:573–82

21. Kleinbaum DG, Kupper LL, Morgenstern H. 1982. *Epidemiologic Research: Principles and Quantitative Methods*. New York: Wiley. 560 pp.

22. Kriebel D, Checkoway H, Pearce N. 2007. Exposure and dose modelling in occupational epidemiology. *Occup. Environ. Med.* 64:492–98
23. Kuha J. 1994. Corrections for exposure measurement error in logistic regression models with an application to nutritional data. *Stat. Med.* 13:1135–48
24. Lagakos SW. 1988. Effects of mismodelling and mismeasuring explanatory variables on tests of their association with a response variable. *Stat. Med.* 7:257–74

25. Lash TL, Fox MP, Fink AK. 2009. *Applying Quantitative Bias Analysis to Epidemiologic Data (Statistics for Biology and Health)*. New York: Springer. 194 pp.

26. Lee LF, Sepanski JH. 1995. Estimation of linear and nonlinear errors-in-variables models using validation data. *J. Am. Stat. Assoc.* 90:130–40
27. Li RF, Weller E, Dockery DW, Neas LM, Spiegelman D. 2006. Association of indoor nitrogen dioxide with respiratory symptoms in children: application of measurement error correction techniques to utilize data from multiple surrogates. *J. Expo. Sci. Environ. Epidemiol.* 16:342–50
28. Lyles R, Kupper L. 1997. A detailed evaluation of adjustment methods for multiplicative measurement error in linear regression with applications in occupational epidemiology. *Biometrics* 53:1008–25
29. Morgan MG, Henrion M, eds. 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. New York: Cambridge Univ. Press. 344 pp.
30. Neas LM, Dockery DW, Ware JH, Spengler JD, Ferris BG, Speizer FE. 1994. Concentration of indoor particulate matter as a determinant of respiratory health in children. *Am. J. Epidemiol.* 139:1088–99
31. Norat T, Bingham S, Ferrari P, Slimani N, Jenab M, et al. 2005. Meat, fish, and colorectal cancer risk: the European Prospective Investigation into Cancer and Nutrition. *J. Natl. Cancer Inst.* 97:906–16
32. Pickles JH. 1982. Air-pollution estimation error and what it does to epidemiological analysis. *Atmos. Environ.* 16:2241–45
33. Prentice RL. 1982. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* 69:331–42

34. Rosner B, Spiegelman D, Willett WC. 1990. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *Am. J. Epidemiol.* 132:734–45

35. Rosner B, Willett WC, Spiegelman D. 1989. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Stat. Med.* 8:1051–69; discussion 1071–73

13. Review of the effects of bias due to mismeasured exposures, and a detailed elucidation of Bayesian sensitivity analysis.

21. Chapter 12 gives an excellent nontechnical introduction to the algebra of misclassification and helps one understand the mathematics of misclassification bias.

25. Includes a chapter on quantitative sensitivity analysis to assess magnitude of bias due to exposure misclassification, and software is publicly available to implement methods.

34 and 35. Multivariate and univariate versions of the popular regression calibration method to correct point and interval estimates in logistic regression models for exposure measurement error. Pubic software is available at http://www.hsph.harvard.edu/faculty/spiegelman/blinplus.html.

36. Seber GAF. 1977. *Linear Regression Analysis.* New York: Wiley. 465 pp.
37. Spiegelman D. 1994. Cost-efficient study designs for relative risk modeling with covariate measurement error. *J. Stat. Plan. Inference* 42:187–208
38. Spiegelman D. 1998. Reliability studies. In *Encyclopedia of Biostatistics*, ed. T Colton, P Armitage, pp. 3771–75. Sussex, England: Wiley
39. Spiegelman D, Carroll RJ, Kipnis V. 2001. Efficient regression calibration for logistic regression in main study/internal validation study designs with an imperfect reference instrument. *Stat. Med.* 20:139–60
40. Spiegelman D, Casella M. 1997. Fully parametric and semiparametric regression models for common events with covariate measurement error in main study/validation study designs. *Biometrics* 53:395–409
41. Spiegelman D, Gray R. 1991. Cost-efficient study designs for binary response data with gaussian covariate measurement error. *Biometrics* 47:851–69
42. Spiegelman D, Hertzmark E. 2005. Easy SAS calculations for risk or prevalence ratios and differences (Invited Brief Commentary). *Am. J. Epidemiol.* 162:199
43. Spiegelman D, McDermott A, Rosner B. 1997. Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *Am. J. Clin. Nutr.* 65(Suppl. 4):1179–86
44. Spiegelman D, Valanis B. 1998. Correcting for bias in relative risk estimates due to exposure measurement error: a case study of occupational exposure to antineoplastics in pharmacists. *Am. J. Public Health* 88:406–12
45. Stram DO. 2005. Designs for studies of personal exposure to air pollution and the impact of measurement error. *J. Toxicol. Environ. Health A* 68:181–87
46. Thomas D, Stram D, Dwyer J. 1993. Exposure measurement error: influence on exposure-disease relationships and methods of correction. *Annu. Rev. Public Health* 14:69–93
47. Thurston SW, Spiegelman D, Ruppert D. 2003. Equivalence of regression calibration methods for main study/external validation study designs. *J. Stat. Plan. Inference* 113:527–39
48. Tosteson TD, Tsiatis AA. 1988. The asymptotic relative efficiency of score tests in a generalized linear model with surrogate covariates. *Biometrika* 75:507–14
49. Tworoger SS, Eliassen AH, Rosner B, Sluss P, Hankinson SE. 2004. Plasma prolactin concentrations and risk of postmenopausal breast cancer. *Cancer Res.* 64:6814–19
50. Valanis BG, Vollmer WM, Labuhn KT, Glass AG. 1993. Acute symptoms associated with antineoplastic drug handling among nurses. *Cancer Nurs.* 16:288–95
51. Van Roosbroeck S, Jacobs J, Janssen NAH, Oldenwening M, Hoek G, Brunekreef B. 2007. Long-term personal exposure to PM2.5, soot and NOx in children attending schools located near busy roads, a validation study. *Atmos. Environ.* 41:3381–94
52. Van Roosbroeck S, Li R, Hoek G, Lebret E, Brunekeef B, Spiegelman D. 2008. Traffic-related outdoor air pollution and respiratory symptoms in children: the impact of adjustment for exposure measurement error. *Epidemiology* 19:409–16
53. Van Roosbroeck S, Wichmann J, Janssen NAH, Hoek G, van Wijnen JH, et al. 2006. Long-term personal exposure to traffic-related air pollution among school children, a validation study. *Sci. Total Environ.* 368:565–73
54. Wacholder S. 1986. Binomial regression in GLIM: estimating risk ratios and risk differences. *Am. J. Epidemiol.* 123:174–84
55. Weller EA, Milton DK, Eisen EA, Spiegelman D. 2007. Regression calibration for logistic regression with multiple surrogates for one exposure. *J. Stat. Plan. Inference* 137:449–61
56. White EJ, Armstrong BK, Saracci R. 2006. *Principles of Exposure Measurement in Epidemiology: Collecting, Evaluating and Improving Measures of Disease Risk Factors*. Oxford/New York: Oxford Univ. Press
57. Xie SX, Wang CY, Prentice RL. 2001. A risk set calibration method for failure time regression by using a covariate reliability sample. *J. R. Stat. Soc. B* 63:855–70

38. A good starting place for literature on the design of validation and reliability studies.

55. Software extension of the regression calibration method to the multiple surrogates setting is available for a single exposure available at http://www.hsph.harvard.edu/faculty/spiegelman/multsurr.htm.

Annual Review of
Public Health

Volume 31, 2010

Contents

Symposium: Public Health Significance of Genomics and Eco-Genetics

Epidemiology and Biostatistics

Social Environment and Behavior

Health in All Policies—The Finnish Initiative: Background, Principles, and Current Issues
Pekka Puska and Timo Ståhl 315

Health Services

Indexes

Errata

An online log of corrections to *Annual Review of Public Health* articles may be found at http://publhealth.annualreviews.org/