Original articles

Scand J Work Environ Health 1996;22:94—101

Maximizing accuracy and precision using individual and grouped exposure assessments

by Noah S Seixas, PhD,[1] Lianne Sheppard, PhD[2]

Seixas NS, Sheppard L. Maximizing accuracy and precision using individual and grouped exposure assessments. *Scand J Work Environ Health* 1996;22:94—101.

Objectives Random errors in exposure data were explored to determine their effect on exposure-response relationships using individual, grouped, or combined (grouped and individual) exposure assessment methods.

Methods Monte Carlo simulations were conducted by generating small "studies" of one hundred subjects divided into four exposure groups. Observed exposure data were generated for each individual using assumed inter- and intraindividual variances and a lognormal distribution. The data were used to calculate the following three estimates of exposure: an individual mean, a group mean, and a hybrid estimate using the James-Stein shrinkage estimator. The exposure estimates were regressed on generated (continuous) "health outcomes," and the regression results were stored and analyzed.

Results Random errors in exposure data resulted in attenuation of the exposure-response relationship when the individual estimates were used, especially when the within-subject variability was high. The attenuation was substantially controlled by the group mean estimate, however, at a cost of decreased precision. The hybrid estimator simultaneously controlled both bias and imprecision in the observed exposure-response function.

Conclusions While estimates of exposure based on individual means may result in attenuation of the exposure-response relationship, grouped estimates may control bias while decreasing precision. Combining individual and group estimates can simultaneously control both types of error. However, further research is required to determine how robust these findings are to different error structures, grouping strategies, exposure-response models, and exposure assessment methods.

Key terms attenuation bias, Berkson error model, James-Stein estimator, measurement error, variance components.

In recent years the importance of accurate exposure assessment to occupational and environmental epidemiology has become widely recognized, and there has been a proliferation of research aimed at improving exposure assessment methods (1—4). Much of the research has been motivated by a recognition of the problems associated with the effects of misclassification or measurement error on exposure-response analyses. In addition, there has been significantly increased attention given to the importance of measurement error issues in epidemiology (5) with a concomitant growth of research on measurement error in biostatistics (6—9). In an overview of measurement error problems in epidemiology, Willett (19) has suggested that "the quantitative assessment of exposure measurement error and correction for its effects . . . is likely to be one of the most fruitful areas of development in epidemiology during the next several years [p 1039]."

While it has become common for occupational exposure assessment experts to refer to the attenuation of effect measures due to randomly misclassified exposure estimates, there has been an inadequate exploration of the specific effects of commonly used occupational or environmental exposure assessment techniques on exposure-response analyses. In particular, while exposure assessment based on measurements taken on individual study subjects is commonly thought of as a "gold standard," feasibility concerns often dictate the use of exposure groups as the basis for assigning individual exposures — implying the introduction of a significant degree of error in the analysis (11). The actual effect of such a grouping strategy on the exposure-response analysis has been inadequately explored.

Epidemiologic exposure assessors have also increasingly recognized the importance of integrating the exposure assessment process with the analysis of health out-

1 Department of Environmental Health, University of Washington, Seattle, Washington, United States.

2 Department of Environmental Health and, Department of Biostatistics, University of Washington, Seattle, Washington, United States.

Reprint requests to: Dr Noah S Seixas, Department of Environmental Health, Box 357234, University of Washington, Seattle, WA 98195, USA.

comes. For instance, it has been argued that a detailed understanding of the toxicologic and pharmacodynamic properties of an agent should be understood to define an appropriate exposure metric for use in an exposure-response analysis (12, 13). An integration of the exposure assessment and exposure-response analysis processes also requires coherence of the statistical methods used (14). These discussions generally imply that one must choose a "best" method or metric and accept the errors associated with that particular choice. An alternative approach is to identify a hybrid method that combines the strengths of each, so as to minimize the overall consequences or errors in the exposure-response analysis.

In this paper, the integration of the exposure assessment and exposure-response processes is discussed in the context of one aspect of exposure assessment, individual and grouped exposure estimation. Linking exposure assessment and exposure-response processes demonstrates the effects of errors associated with each method through computer simulation.

Methods

The results of individual and group exposure assessment methods for observed exposure-response relationships were explored with the use of a simple exposure-response study with Monte-Carlo simulation. The scenario assumed a small number of exposure groups in a study population with an equal number of subjects in each group and an equal number of exposure measurements taken on each individual.

Simulations were conducted with four exposure groups with a total of 100 subjects in each realization of a "study." All the exposure values were assigned on the assumption that the data were lognormally distributed, and the analysis was conducted on the natural logarithms of the original values. The group geometric means, μ_g, were assigned as fixed values of 2.7, 7.4, 20.1, and 54.6 (exponents of 1, 2, 3, and 4). "True" individual exposures (on a log scale), y_{gi}, were generated from a normal distribution, given the log of the group geometric mean and an assumed interindividual variance, as represented by the interindividual geometric standard deviation (GSD_B); that is, $y_{gi} \sim N\{\ln(\mu_g),[\ln(GSD_B)]^2\}$. An "observed" health outcome was then generated using a linear model:

$$w_{gi} = \alpha + \beta(y_{gi}) + \varepsilon_{gi},$$

where w_{gi} is the generated health outcome for individual i in group g. β is the exposure-response regression coefficient, set to 1 for generating the observed outcomes, and the error variance, ε_{gi}, is normally distributed, the mean being 0 and the variance being equal to 0.01.

A set of two "observed" exposure measurements, z_{gij}, was then generated for each individual using the subjects "true" exposure and a defined intraindividual variance: $z_{gij} \sim N\{y_{gi},[\ln(GSD_W)]^2\}$. From this set of generated raw data, the observed individual ($\bar{z}_{gi\cdot}$) and group ($\bar{z}_{g\cdot\cdot}$) mean exposures were calculated and regressed (in separate models) on the observed outcome. The regression coefficients were stored and the process, using the same input parameters, was repeated 300 times to ensure reasonably stable estimates. The mean regression coefficient and its standard error were then calculated and reported.

The variance components were chosen to reflect the range that might be expected in typical occupational environments with the total variability, as expressed by the total geometric standard deviation (GSD_T) of 1.5, 2.5, 3.5, and 4.5. The total variability was distributed between the within person (GSD_W) and between person (GSD_B) variance components according to the following:

$$[\ln(GSD_T)]^2 = [\ln(GSD_B)]^2 + [\ln(GSD_W)]^2 .$$

Values of GSD_B and GSD_W of 1.33, 1.91, 2.66, and 3.89 were adopted to satisfy this relationship. To limit the GSD_T to 4.5, no combinations of GSD components of 2.66 and 3.89 or 3.89 and 3.89 were used.

Results

The results of the simulation with individual and grouped exposure estimates are given in table 1. With the individual mean for the exposure variable, the attenuation of the regression coefficient is clearly observed, especially when the intraindividual variance is large and the interindividual variance is relatively small. The standard error of the coefficient follows a similar trend, generally becoming larger as the attenuation bias increases. In contrast, use of the group mean as the predictor substantially eliminated attenuation bias. Increasing the interindividual variability resulted in substantially increased uncertainty (larger standard errors), while the standard error was not substantially affected by an increase in intraindividual variability. Group mean exposure estimates produced unbiased exposure-response estimates with lower standard errors than the individual means in situations where the intraindividual variability was relatively high and the interindividual variance was low.

For a visualization of these results, the data and estimates from a single iteration, using inter- and intraindividual GSD values of 2.66, were plotted and are shown in figures 1A and 1B. Figure 1A shows the 100 individual means plotted against the outcome with the "expected" regression line (solid line) of the slope equal to 1 and the least squares fit line (broken line), demonstrating

Table 1. Simulation results (mean coefficient and standard error) for 100 persons distributed into four exposure groups over a range of inter- and intraindividual variances using individual and group means as the exposure metric.[a] (GSD_W = intraindividual geometric standard deviation, GSD_B = interindividual geometric standard deviation, GSD_T = total geometric standard deviation)

GSD_W	Individual mean								Group mean							
	GSD_B 1.33		GSD_B 1.91		GSD_B 2.66		GSD_B 3.89		GSD_B 1.33		GSD_B 1.91		GSD_B 2.66		GSD_B 3.89	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1.33	0.97	0.02	0.98	0.02	0.98	0.02	0.99	0.01	1.00	0.03	1.00	0.06	1.00	0.09	1.00	0.12
1.91	0.87	0.04	0.89	0.03	0.91	0.03	0.94	0.03	1.00	0.03	1.01	0.06	0.99	0.09	1.00	0.12
2.66	0.74	0.05	0.79	0.04	0.82	0.04	[b]	.	1.00	0.03	1.00	0.06	0.99	0.09	[b]	.
3.89	0.59	0.05	0.65	0.05	[b]	.	[b]	.	0.98	0.03	0.99	0.06	[b]	.	[b]	.

[a] The simulation parameters were four groups of 25 individuals each with group geometric means of 2.7, 7.4, 20.1, and 54.6 and two observed exposure measurements each. The results given are the mean regression coefficients and standard errors (SE) generated by 300 iterations.
[b] These variance components were not utilized so that the GSD_T could be constrained to < 4.5.

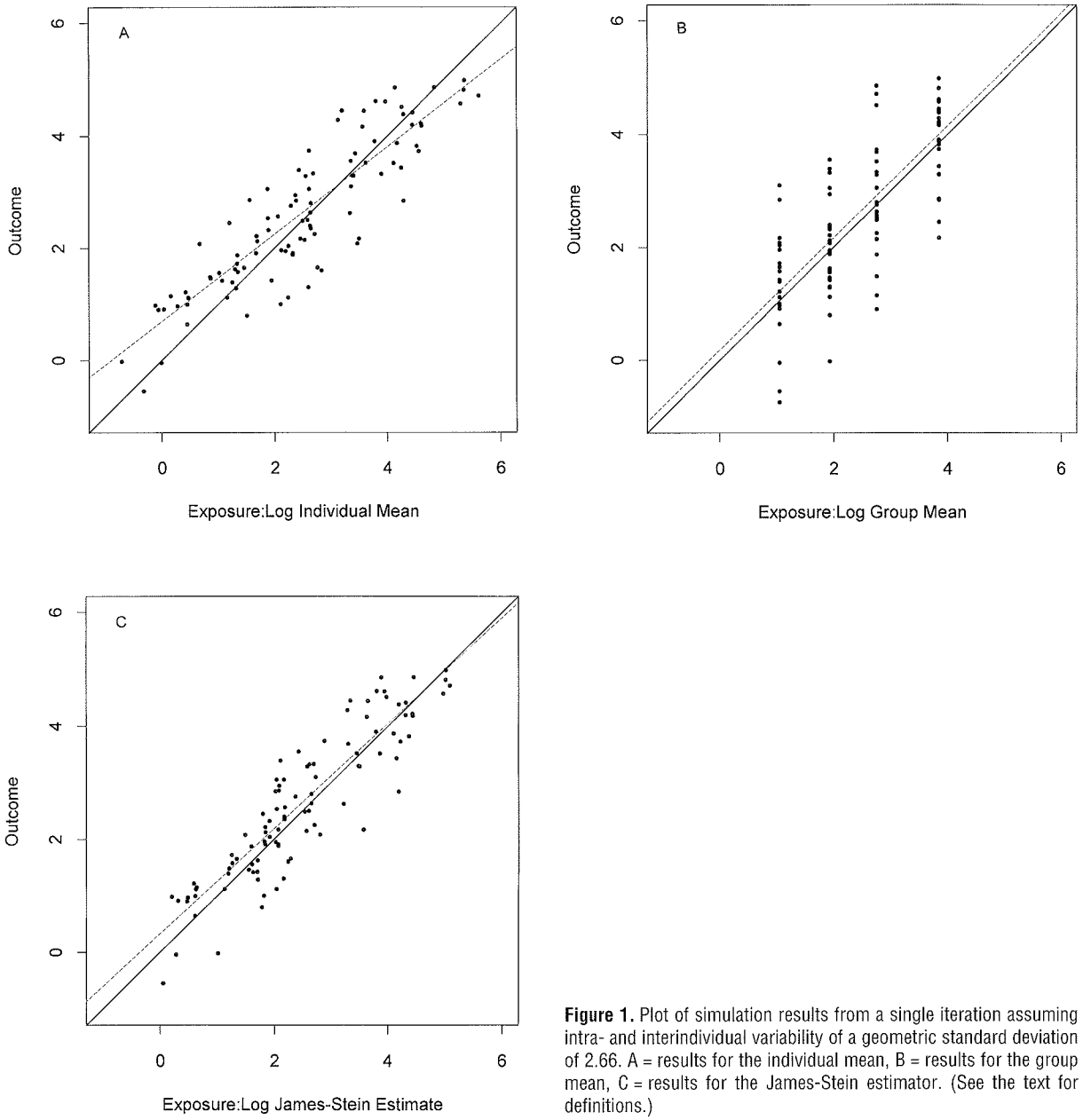

Figure 1. Plot of simulation results from a single iteration assuming intra- and interindividual variability of a geometric standard deviation of 2.66. A = results for the individual mean, B = results for the group mean, C = results for the James-Stein estimator. (See the text for definitions.)

the attenuation of the slope due to random error. Figure 1B shows the same 100 individuals plotted according to their group means. Though the regression line of the grouped data is very close to the expected value, it is easy to imagine how increased levels of interindividual variability, resulting in a larger spread of the data along the y-axis, would lead to instability in the resulting slope parameter.

These simulations demonstrate that when there is relatively little interindividual variability within groups, the group mean performs well by eliminating attenuation bias at only a small cost of increased variance. In fact, when the GSD_B is low and the intraindividual variability is large, the group mean actually out-performs the individual mean with respect to both bias and uncertainty. However, when the GSD_B is large, the attenuation of the individual mean bias is relatively small and the increased variability associated with the use of the group mean is not compelling.

In summary, there is a trade-off between bias and variance in the choice of the individual or group mean for exposure assessment strategies. Generally, the individual mean will be associated with bias in the regression coefficient, whereas the group mean substantially eliminates this bias. However, the elimination of bias comes at a cost of increased uncertainty, which under some circumstances may be a very poor trade-off. These observations led us to consider whether there is some way to exploit both the stability associated with the spread of individual means and the decrease in bias associated with grouping individuals.

It is useful to consider the individual and group means as two plausible estimators of individual exposure, each contributing different but important information. The individual mean is an unbiased estimate of true individual exposure because, in random sampling, each exposure measurement is an unbiased estimate of true exposure. But the individual estimate is also imprecise because there is only a small number of samples per individual. Use of the individual exposure estimate leads to attenuation of the exposure-response slope, or regression coefficient, β. On the other hand, the group mean has greater precision due to its combining information across all group members, but it is a biased estimate of each individual's true exposure level. Use of the group mean exposure estimate in a regression model leads to an unbiased regression coefficient. The slope parameter is estimated with greater uncertainty because there is less variation in the predictor relative to the outcome.

Ideally, an estimator of exposure would improve both the precision and the bias of the observed exposure-response regression coefficient. We propose a weighted mean of the two simpler quantities:

$$\Theta_{gi} = (B_g)\bar{z}_{g\bullet\bullet} + (1 - B_g)\bar{z}_{gi\bullet},$$

where $\bar{z}_{g\bullet\bullet}$ is the group mean exposure and $\bar{z}_{gi\bullet}$ is the individual mean. B_g is the weighting factor that determines the optimal combination of the group and individual means. This estimator, Θ, is called the James-Stein shrinkage estimator, originally described by James & Stein (15) and further discussed by Efron & Morris (16) among others. Though originally developed to improve multivariate estimates of population parameters, it has more recently been shown to be effective in reducing measurement error bias in linear and nonlinear regression, on the assumption of a Gaussian error distribution (17). We estimate B_g as

$$\hat{B}_g = \min\left[1, \frac{\hat{\delta}_g J^{-1}}{\hat{S}_g}\right],$$

where

$$\hat{\delta}_g = \frac{\sum_{i=1}^{n}\sum_{j=1}^{J}(z_{gij} - \bar{z}_{gi\bullet})^2}{n(J-1)+2}$$

and

$$\hat{S}_g = \frac{\sum_{i=1}^{n}(\bar{z}_{gi\bullet} - \bar{z}_{g\bullet\bullet})^2}{n-3}$$

for z_{gij} measurements taken on $i=1, \ldots, n$ workers in $g=1, \ldots, G$ groups over $j=1, \ldots, J$ repetitions (or days). This weighting factor estimates

$$\frac{(\ln(GSD_W))^2 J^{-1}}{[\ln(GSD_B)]^2 + [\ln(GSD_W)]^2 J^{-1}}$$

for a fixed number of repeated measurements on each individual. As the intraindividual variability increases relative to the interindividual variability (and the expected measurement error attenuation bias increases), more weight is put on the group mean (in equation 1); the bias is controlled. However, when the intragroup variance is small in comparison with the intergroup variance, then more weight is placed on the individual mean, the spread of the exposure variable is maximized, and the uncertainty in the exposure-response relationship is reduced.

The James-Stein estimator has been described previously for a single analytic group [see, eg, Whittemore (17)], that is, where there is a single group mean (the population mean) rather than several group-specific means. We have generalized the estimator to be group-specific under the assumption that the exposure groups contain some meaningful information about individual exposures. Thus the shrinkage given by the James-Stein estimator is towards the group mean rather than towards the population mean. This is consistent with our focus on the comparison between individual and group mean exposures in exposure-response models. We plan further

methodological development to clarify the benefit of an overall versus a group-specific shrinkage estimator.

The simulation conducted previously using the individual or group means was re-run with the James-Stein estimator as the exposure metric; the results are provided in table 2. As predicted, the attenuation bias was substantially less than it was for the individual mean, and the standard errors of the estimates were likewise reduced from the group mean case. The dispersion of individual exposure estimates having been reduced with the use of the group means, a better estimate of the regression line was obtained, as shown in figure 1C. Thus it appears that if strength is borrowed from both the group and individual mean estimates to create a combined exposure estimate, problems associated with the use of either estimator alone in a regression model may be reduced.

Table 2. Simulation results using the James-Stein estimator as the exposure metric.[a] (GSD_W = intraindividual geometric standard deviation, GSD_B = interindividual geometric standard deviation, GSD_T = total geometric standard deviation)

GSD_W	GSD_B							
	1.33		1.91		2.66		3.89	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
1.33	1.00	0.02	1.00	0.02	1.00	0.02	1.00	0.01
1.91	0.98	0.03	0.99	0.03	0.98	0.03	0.99	0.03
2.66	0.95	0.03	0.96	0.04	0.96	0.04	.[b]	.
3.89	0.91	0.04	0.92	0.05	.[b]	.	.[b]	.

[a] The simulation parameters were 100 individuals in four exposure groups of 25 each. The results given are the mean regression coefficient and standard errors (SE) generated by 300 iterations.
[b] These variance components were not utilized so that the GSD_T could be constrained to < 4.5.

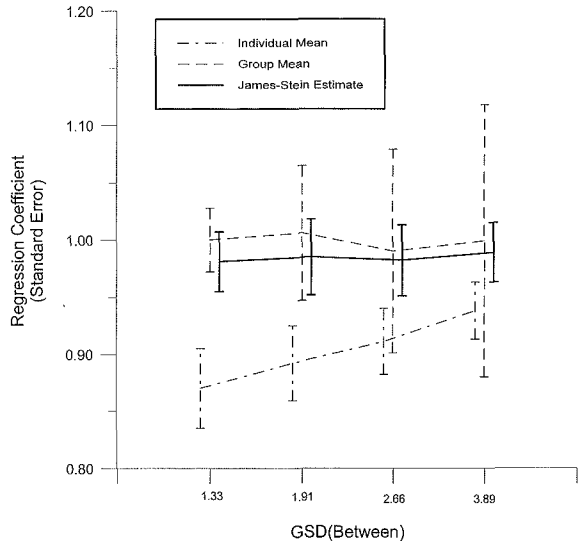

Figure 2. Results of simulations using individual, group, and James-Stein estimates of exposure on the assumption of a geometric standard deviation for intraindivisual variance components (GSD_W) equal to 1.91 and varying levels of the geometric standard deviation for interindividual variance components (GSD_B). The results are shown as the mean regression coefficients and their associated standard errors (error bars).

A visual comparison of the group, individual, and James-Stein estimator results from tables 1 and 2 with the intraindividual variability of GSD_W equal to 1.91 is presented in figure 2. In this display, the bias associated with the individual mean and the large standard errors associated with the group mean are clearly observed, whereas the James-Stein estimator substantially reduces the bias while maintaining reasonably high precision.

Discussion

Errors in the quantification of exposure, including both bias and imprecision, are a primary limiting factor in many occupational epidemiology studies because of their effects on exposure-response analyses. A simple systematic error or bias in an exposure metric (on an additive scale) produces a shift in the observed exposure-response relationship of the same magnitude as the bias. As a consequence of such a shift, the degree of disease predicted at a particular level of exposure may be incorrect, and more or less disease than expected may occur at the stated level of exposure, depending on the direction of the bias. A bias of this type in an exposure-response relationship might be used to produce either an unnecessarily strict, or inadequately protective, exposure guideline. However, in the case of an additive bias uncorrelated with the outcome, the slope of the exposure-response function is unaffected, and the increment of increased or decreased disease expected for a given change in exposure levels (the slope of the regression line) would be accurately estimated.

A measured or estimated exposure metric may also have a substantial degree of random error associated with it. Random errors may arise for a variety of reasons depending on the data available, the methods used to summarize the data, and the metric adopted to express each individual's exposure. For instance, it is well known that most airborne occupational exposures vary from day to day with a pattern that approximates a lognormal distribution with geometric standard deviations of four or even higher (3, 18). As a result of this high environmental variability, the long-term mean exposure estimated with a small number of measurements will have a large standard error, that is, it is estimated with a high degree of uncertainty.

According to classical measurement error theory, such random error in exposure results in an exposure-response relationship that is generally attenuated toward the null (7). Like a systematically biased exposure metric, attenuation bias from random errors will usually result in an inaccurately predicted degree of disease, given a specified degree of exposure. More importantly, in the

context of etiologic epidemiology, measurement error bias may obscure the recognition of an underlying causal association. The degree to which random error biases the exposure-response function is determined by the degree of imprecision in the individual's exposure metric (the intraindividual variance) and the variance of exposure across the studied population (the interindividual variance), as was demonstrated in the simulations.

The relationship between the true underlying exposure-response function and the observed regression coefficient, given classical measurement error, may be described by:

$$\beta = b(1 + \lambda),$$

$$\lambda = \frac{\sigma_W^2}{\sigma_B^2},$$

where β is the true regression coefficient and b is the observed coefficient given intra- and interindividual variances in exposure of σ^2_w and σ^2_B, respectively. Thus, if σ^2_w is of the same magnitude as σ^2_B , λ equals 1 and the observed regression coefficient would be half that of the true exposure-response relationship. Note that σ^2_B here is the variation in exposures across the whole population, and not within an exposure group as in the simulations. As the precision of the individual's exposure metric becomes small compared with the spread of the exposures across the population, the observed coefficient approaches the true relationship.

For classical measurement error, the uncertainty in an individual's measurement results in a spreading or overdispersion of the exposure values along the independent variable in comparison with the outcome, and, as a result, the regression line is attenuated. If exposure is collapsed to a group mean, and assuming a linear exposure-response model, the errors associated with each individual's exposure are averaged across the group and, for the group as a whole, are close to 0. As a result of this grouping, the overdispersion in exposure is controlled, although some of the information about individual true exposures is also eliminated. While exposure is still mismeasured under these conditions, attenuation of the regression coefficient in the exposure-response analysis is substantially controlled, as demonstrated in this paper. The control of attenuation due to measurement error with appropriate grouping is due to the much smaller error variance of the group mean relative to the individual mean. It may also be approximately described by the Berkson error model (7). Under the Berkson model, the bias in the regression coefficient is reduced, while the uncertainty associated with the estimate may increase. Thus, as demonstrated in the current simulations, a tradeoff is implied by using a grouped exposure assessment: while grouping reduces bias in the exposure-response model, it also increases uncertainty. With either individual or group exposure assessment approaches, the likelihood of observing a statistically significant exposure-response relationship, where one truly exists, is therefore reduced.

In most occupational epidemiology studies, individual estimates are either unavailable or have low precision because there are insufficient data available on each subject to estimate individual exposures reliably. As a result, exposure groups are assumed to be predictive of exposure for all group members. Groups may be defined on the basis of worksite, job title, or department but may also be defined by a more complex set of modeled exposure determinants. Quantitation is done on the basis of the group, and some relevant parameter of the group's exposure distribution, usually the arithmetic mean, is assigned to each individual in that group. The validity of this approach rests on the assumption that the exposure data available for the group are truly representative. Generally, representativeness is obtained by taking a random sample of individuals in the group, and days over the period of interest, although other more efficient strategies could be adopted to obtain a representative estimate for the group.

Grouped exposure assessment has been used by epidemiologists for a long time (19). More recently, the efficiency of grouping for exposure assessment has been recognized (20) and widely accepted by the industrial hygiene community as the basis for efficient strategies (21). The term homogeneous exposure group (HEG) has been widely adopted, implying that not only is grouping a more efficient strategy for assessment, but that individuals within an identified group have very similar day-to-day exposure distributions and essentially the same expected or long-term mean exposure.

The homogeneity assumption has been substantially challenged through the analysis of intra- and interindividual variance components from a set of 165 exposure groups defined in published studies (11, 22). Each dataset in these analyses included repeated measurements of exposure on each study subject, and the results demonstrated very high interindividual variances within many groups. This analysis concluded that most homogeneous exposure groups are actually inhomogeneous, with a very wide spread of long-term individual means. About 30% of the groups had over 10-fold differences between the lower 2.5th and 97.5th percentiles of the individual means. Thus the use of the group mean exposure in these studies implies a high degree of error in the assigned exposures in comparison with the true individual exposures.

The view that we must choose either grouped or individual assessment is very limited. The use of exposure groups implies that we believe that individuals belonging to that group have something in common in relation to exposure — that is, group membership is a

useful, if imperfect, predictor of exposure. If we ignore the information derived from group membership itself, we have lost valuable information. However, we also know that individuals within groups differ from each other — that there is individual-specific information that distinguishes individuals. By utilizing both the information that distinguishes the individual from the group and information about the group that makes it an identifiable unit, we may be able to maximize the use of all available information and thereby reduce the effects of measurement error in terms of both accuracy and precision.

In the particular context developed in this presentation, in which group and individual means were calculated from data assumed to be randomly collected from each group member, the James-Stein shrinkage estimator was shown to work well in simulation. A different variance-weighted estimator has been presented to attempt to resolve the on-going debate about the apparent association between electromagnetic fields and childhood leukemia (23). Exposure represented by a simple categorization of wiring configuration near the residence has been associated with childhood leukemia in several studies, while individual exposure measurements have generally not been predictive of the outcome. In this study, individual measurements of residential electromagnetic fields were combined with the estimated mean field-strength predicted from a model developed from wiring configuration and other factors. The estimator used a variance-weighted combination similar to the James-Stein estimator; however it relied on information from the predictive model rather than just the grouping variable. In this situation, no improvement in the exposure-response relationship was observed after the two sets of information were combined, and the authors concluded that the long-term mean exposure intensity predicted by their model, and estimated by the measurement, was probably not the appropriate summary exposure metric.

The presented simulations contain some assumptions. First, we have assumed that the exposure distributions are lognormal and that the exposure-response analysis is conducted on the log-transformed exposure and outcome data. This approach is by no means uniformly adopted, but is not unreasonable since it implies multiplicative relations between covariates and supports the use of hypothesis tests which rely on normally distributed errors. The extension of this analysis to other data distributions and nonlinear models requires further development.

Second, we have considered the group geometric mean values as fixed effects, rather than as randomly distributed parameters with an intergroup variance component. In future development we plan to consider a particular study setting as only an example of the universe of study situations and to extend the analysis to address the intergroup variance, as well as the interindividual and intraindividual variance components. This approach will allow a more generalizable conclusion concerning the three components of variance. However, to demonstrate the effect of the three assessment strategies addressed, the use of fixed group means as a simplification does not seem unreasonable. In effect, the fixed effects imply that a given industrial setting is studied and the true group exposure levels are stationary or fixed quantities that are determined prior to the study.

Third, although our study was confined to simulations, the study was designed to represent typical cross-sectional investigations in which a small number of job groups are observed and a limited number of measurements are taken on each subject. For instance, a study of pulmonary function changes over a workshift and workweek might involve exposure measurements taken on each subject on the Monday and Friday of spirometric testing. These measurements could also then be used on a group basis to determine the long-term mean exposure of individuals performing that job.

Finally, and perhaps most importantly, we have assumed that the data gathered are truly representative of exposure. While this assumption is integral to our simulations, the realities of field data collection frequently make the goal of obtaining representative random samples elusive. In fact, nonrepresentative sampling may introduce biases in the exposure variables which would add another potentially very significant type of error to these analyses.

Although our ability to address this question is limited by the constraints of the developed simulations, what do the current results suggest in terms of how such grouping should be conducted? While the reduction of bias demonstrated with a grouping method suggests that defining the groups as inclusively as possible (a large number of individuals in a small number of groups) would be advantageous, the groups should be defined to preserve the largest possible spread of exposures between groups. Broadening a group's definition to include a larger number of individuals involves both increasing precision (due to more data on the group and consequently lower error variance) and decreasing precision due to a larger group exposure variance (assuming that the group becomes more heterogeneous as it increases in size). In addition, defining a larger group will tend to move the group means toward the population mean and thus narrow the spread of exposure values used across the population and increase the uncertainty with which the exposure-response relationship can be estimated. Thus understanding the effects of alternative grouping strategies will depend on the analysis of intra- and interindividual variance components under the different strategies; sampling campaigns should be conducted to allow for these analyses. The efficient design of grouping strategies must also consider alternative exposure distribu-

tions, unbalanced exposure sampling designs, nonlinear exposure-response relations, and group misclassification rates. Additional considerations must be made if there is residual confounding introduced by unmeasured, mismeasured, or simply confounding covariates. Simulation studies to address these issues are now being planned.

Extension of the concept of combining different types or levels of information to achieve an improved exposure metric in other contexts will require modified approaches. For instance, subjective estimates derived for time periods in which there were no industrial hygiene measurements taken might be combined effectively with quantitative measurements from other periods, if the nature of the errors associated with both information sources can be understood sufficiently.

Given the substantial limitations of many types of available exposure information, methods which combine the strengths of several types of data into the exposure assessment process — in ways which reduce the effects of the errors of each — are required for continuing progress in occupational epidemiology. The proposed method, combining individual and grouped mean exposure levels through the use of the James-Stein estimator, is only one limited application of this concept. Combining alternative types of data through an understanding of their error structures and explicit linking of the exposure assessment and exposure-response analyses will help derive strength from each source of exposure information. In this manner, the effects of random errors can be minimized, and our understanding of occupational health hazards will continue to advance.

Acknowledgments

The authors thank Larry Engel for his expert programming assistance and Tom Robins for his careful review of an early draft.

References

1. Checkoway H, Pearce NE, Crawford-Brown DJ. Research methods in occupational epidemiology. New York (NY): Oxford University Press, 1989.
2. Checkoway H, Seixas NS, Demers PA. The influence of occupational exposure assessment on epidemiologic inferences. Occup Hyg. In press.
3. Rappaport SM, Smith TJ. Exposure assessment for epidemiology and hazard control, industrial hygiene science series. Chelsea (MI): Lewis Publishers, 1991:313.
4. Stewart PA, Herrick RF. Issues in performing retrospective exposure assessment. Appl Occup Environ Hyg 1991;6:421—7.
5. Armstrong BK, White E, Saracci R. Principles of exposure measurement in epidemiology. Oxford: Oxford University Press, 1992.
6. Fuller WA. Measurement error models. New York (NY): Wiley, 1987.
7. Thomas D, Stram D, Dwyer J. Exposure measurement error: influence on exposure-disease relationships and methods of correction. Ann Rev Public Health 1993;14:69—93.
8. Byar DP, Gail MH. Errors-in-variables workshop. In: Statistics in medicine. New York (NY): John Wiley and Sons Ltd, 1989:8.
9. Carroll RJ, Stefanski LA, Ruppert D. Nonlinear measurement error models. New York (NY): Chapman and Hall, 1995.
10. Willett W. An overview of issues related to the correction of non-differential exposure measurement error in epidemiologic studies. Stat Med 1989;8:1031—40.
11. Rappaport SM, Kromhout H, Symanski E. Variation of exposure between workers in homogeneous exposure groups. Am Ind Hyg Assoc J 1993;54:654—62.
12. Kriebel D. The dosimetric model in occupational and environmental epidemiology. Occup Hyg 1994;1:55—68.
13. Salvan A, Stayner L, Steenland K, Smith R. Selecting and exposure lag period. Epidemiology 1995;6:387—90.
14. Seixas NS, Robins TG, Moulton LH. The use of geometric and arithmetic mean exposures in occupational epidemiology. Am J Ind Med 1988;14:465—77.
15. James W, Stein C. Estimation with quadratic loss. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley (CA): University of California Press, 1961:361—80.
16. Efron B, Morris C. Limiting the risk of Bayes and empirical Bays estimators — part II: the empirical Bayes case. J Am Stat Assoc 1972;67:130—9.
17. Whittemore AS. Error-in-variables regression using Stein estimates. Am Stat 1989;43:226—8.
18. Buringh E, Lanting R. Exposure variability in the workplace: its implications for the assessment of compliance. Am Ind Hyg Assoc J 1991;52:6—13.
19. Ashford JR. The design of a long-term sampling programme to measure the hazard associated with an industrial environment. Royal Stat Soc J 1958;121:333—47.
20. Corn M, Esmen NA. Workplace exposure zones for classification of employee exposure to physical and chemical agents. Am Ind Hyg Assoc J 1979;40:47—57.
21. Hawkins NC, Norwood SK, Rock JC. A strategy for occupational exposure assessment. Akron (OH): American Industrial Hygiene Association, 1991.
22. Kromhout H, Loomis DP, Mihlan GJ, Peipins LA, Kleckner RC, Iriye R, et al. Assessment and grouping of occupational magnetic field exposure in five electric utility companies. Scand J Work Environ Health 1995;21:43—50.
23. Thomas D, Bowman J, Jiang L, Jiang F, Peters J. Residential magnetic fields predicted from wiring configurations: relationships to childhood leukemia. Bioelectromagnetics. In press.

Received for publication: 6 April 1995