

Effects of approximation in exposure assessments on estimates of exposure-response relationships

by Benedict G Armstrong, MSc, David Oakes, PhD¹

ARMSTRONG BG, OAKES D. Effects of approximation in exposure assessments on estimates of exposure-response relationships. *Scand j work environ health* 8 (1982): suppl 1, 20-23. Information on exposure in cohort studies of occupational groups is often very incomplete. If assessments of exposure level are attempted, they are inevitably subject to random error. Even when unbiased, such measurement error can give rise to bias in an estimate of an exposure-response relationship. Since practical techniques exist to account for this effect only in certain situations where response may be measured on a continuous scale, a technique suitable for mortality studies needs developing. In the present report a procedure is described for reducing the bias due to random error in the estimated exposure-response relationship from a cohort mortality study. This procedure is based on a mechanism for adjusting exposure assessments, taking into account the distribution of exposures, and the independently estimated distribution of measurement errors. The initial bias and its reduction through analysis is shown for some simulated data sets. The results of applying this technique to a cohort mortality study of asbestos workers are presented.

Key terms: asbestos, dose-response, measurement error, occupational mortality.

Most occupational epidemiology is concerned with providing evidence on *qualitative* questions of whether particular exposures or work conditions cause adverse health effects. In a number of cases, however, an adverse effect is known to exist, but total elimination of exposure is not possible. Then hygiene limits must be set from *quantitative* evidence of the exposure-response relationship.

Before such relationships can be estimated exposures experienced by study subjects must be assessed. Occasionally, for example in some industries entailing exposure to radiation, assessments of ex-

posure to individuals may be available from personal dosimeters. Usually, workers' exposures must be deduced from their work histories and estimates of exposure levels for each job. Exposure may be aggregated over time in particular ways to suit hypotheses of latency period, etc, but the simple measure of cumulative exposure is often used.

All exposure estimates are approximate. Consistent overestimation or underestimation will clearly bias the estimate of the exposure-response relationship. The situation considered in this communication, however, is when, on the average, exposures are neither underestimated nor overestimated, so that over the whole study there is *random unbiased* error in the exposure assessments. This problem is akin to that of missclassification in 2×2 tables (2), a situation which is applicable where exposure and response are measured dichotomously.

¹ TUC Centenary Institute of Occupational Health, London School of Hygiene and Tropical Medicine, London, England.

Reprint requests to: Mr BG Armstrong, TUC Centenary Institute of Occupational Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, England.

Known techniques

It has long been known that random unbiased error in measuring an exposure variable tends to bias the slope of an exposure-response line towards zero. When response is measured on a continuous scale, techniques are available in many situations to adjust for this trend if the measurement error variance is known (3).

Say, we have for each subject an unobserved true exposure X , with mean μ_x and variance σ_x^2 , an observed exposure Z which includes random error ϵ with mean 0 and variance σ_ϵ^2 so that

$$Z = X + \epsilon,$$

a response Y with, for a given Z ,

$$E(Y|X) = \alpha + \beta X. \quad (a)$$

If Y is measured on a continuous scale and the errors, ϵ , are not dependent on X , then the usual least squares estimate b of β using Z instead of X will be biased with

$$E(b) = R\beta : R = \frac{\text{Var}(X)}{\text{Var}(Z)} = \frac{\text{Var}(Z) - \sigma_\epsilon^2}{\text{Var}(Z)}. \quad (b)$$

If we know σ_ϵ^2 , we may estimate R , since $\text{Var}(Z)$ may be estimated from the data. An unbiased estimate b/R of β , can then be computed.

This method of accounting for measurement error, known as correction for attenuation, requires that the response Y be continuous, and that ϵ be independent of X . Thus dichotomous response or measurements of different precision cannot be properly accommodated. Sometimes, however, the relationship in model b gives a rough idea of whether measurement error is likely to give rise to serious bias. For example, if the standard deviation of true exposures is double that of measurement error, the resulting estimate of slope b would be, on the average, $4/(4+1) = 80\%$ of the true slope.

In some situations an estimate of the exposure-response relationship accounting for measurement error is still required, but the correction for the attenuation approach is not appropriate. In theory it is possible to find a maximum likelihood estimate of the true exposure-response slope if parametric forms are assumed for the distributions of the errors and true exposures, and if enough information is available to estimate the parameters of these distributions

(8). The computational burden makes this procedure impractical for most studies.

A proposed method

Suppose that we have X , Z , ϵ , and Y as before, except that Y is not assumed continuous, and ϵ need not be independent of X . If the joint distribution of X and ϵ were known, it would be possible, using Bayes Law, to compute the mean value, $E(X|Z)$, of possible true exposures, X , for each observed exposure, Z . Following standard usage, we will call this the posterior mean of X given Z . If we assume, as before, that

$$E(Y|X) = \alpha + \beta X, \quad (c)$$

then

$$E(Y|Z) = E(E(Y|X)|Z) = \alpha + \beta E(X|Z). \quad (d)$$

Notice that equation d has the same form as equation c, with X replaced by $E(X|Z)$, its posterior mean given Z . This result suggests that using the posterior means of X given Z in place of the observed exposures Z in whatever analysis is used to estimate β may give a less biased estimate.

The distributions of ϵ and X are not generally known. However, if data are available to estimate the distribution of ϵ (or if it is assumed to be known), then the posterior means may still be estimated from the observed distribution of Z , with an "Empirical Bayes" method (6).

The effectiveness of this method in reducing the bias due to random measurement error will vary according to the situation in which it is applied. It is simpler to carry out, but it will not in general perform as well as the maximum likelihood estimate mentioned at the end of the preceding section.

Nonlinear exposure-response relationships may be assumed. In these cases modifications which make the proposed method applicable will generally be possible.

Simulations

To illustrate the effect of some typical errors in exposure measurement, and its reduction using the method proposed, analyses were performed on simulated mortality data. The distribution of true exposures over the population at risk was

Table 1. Effect of simulated measurement error.

Error standard deviation (log scale)	True slope	Mean estimates of β^a					
		True exposure X		Approximate exposure Z		Posterior mean $E(X Z)$	
0.5	0.1	0.112	(0.007)	0.0793	(0.006)	0.104	(0.008)
0.5	1	1.02	(0.07)	0.548	(0.03)	0.902	(0.08)
1	0.1	0.105	(0.006)	0.0362	(0.002)	0.0865	(0.007)
1	1	1.05	(0.1)	0.185	(0.01)	0.685	(0.07)

^a Standard error in parentheses.

taken to be log-normal, as is often found in practice (9). The relative risk experienced by the subjects (compared to non-exposed workers) was assumed to be a linear function of exposure:

$$r = 1 + \beta X. \quad (e)$$

The simulated exposures were grouped into sets of equal size and one death was chosen from each set randomly, but taking into account the different risks computed from equation e. The simulated exposures, X , were then subjected to unbiased random error ε , again taken to be log-normal. Values of the resulting approximate exposures Z were thus computed for each X with:

$$Z = X\varepsilon; \quad E(\varepsilon) = 1.$$

This simulation process produced case-referent sets from which β in equation e could be estimated with the appropriate likelihood method for matched case-referent sets (1). Three estimates of β were made, using (i) true exposures X , (ii) approximate exposures Z , and (iii) posterior means $E(X|Z)$, calculated as explained in the section presenting the proposed method, using known values of the parameters of the distribution of ε and X .

The results of the simulations are shown in table 1 for a distribution of true values X with mean $(\text{LOG}(X)) = 1$ and SD $(\text{log}(X)) = 1$. Fifty simulations of 250 sets of one case and four referents were used to compute 50 values of the three estimates of β . The means of these estimates and their standard errors are presented in the table.

The effect of measurement error is noticeable even with the smaller error, and substantial with the larger error. The effect of the greater spread of risk in the population, and hence greater β , is seen to

increase the bias due to measurement error.

Using the posterior means to estimate β appears more effective in reducing bias where the error, and the spread of risk in the population, is smaller. Where the spread of risk in the population is very large, giving a very long tail to the distribution of exposures, the use of posterior means in place of the observed exposures can increase the bias. Work is in progress to determine why this is the case, and to develop an improved method which avoids this problem.

While these simulations were case-referent sets, the same basic relationships may be expected to apply in cohort studies.

An example: Quebec chrysotile workers

The mortality of a cohort of 11,379 Quebec chrysotile asbestos miners and millers has been studied extensively (7). Work histories were recorded for each subject, and some 5,783 jobs identified. For each job-year an average dust concentration was estimated in millions of particles per cubic foot (mpcf).² Estimates were based on over 4,000 midjet impinger dust counts, but were inevitably subject to considerable approximation, particularly for the earlier years, when few measurements were available and conditions were known to be worse. Cumulative dust exposures (mpcf-years) were then computed for each worker.

Liddell et al (5) argued that case-referent sets from within the cohort could be used efficiently to estimate an exposure-response relationship. Using this procedure,

² 1 mpcf = 35.31 million particles/m³.

we have taken referents for the 245 lung cancer deaths matched for age and smoking. Assuming model e, the maximum likelihood estimate of slope β , computed from the observed cumulative exposures, is 0.00131 (0.0006).

The posterior means, as defined earlier, for workers' cumulative dust exposure are found through summing the posterior means of the exposures accumulated from each job held. Thus the requirement is to find posterior means of the assessments of each exposure for each job with the methods presented in the Proposed Method section. The distribution of errors was assumed to be log-normal. Error variance for the period for which dust measurements were available was estimated to be 0.3 on the log scale, with the use of replicated midjet impinger measurements from the 4,000 used to assess exposures. From discussions with hygienists, error variance was taken to be 0.6 for the period 1930–1940, and 0.9 for the time before 1930. Since true levels of exposure would depend on the year, the mine, and the departments within which the job was carried out, the distribution of exposures was taken to be log-normal about a linear structure reflecting year, mine, and department effects.

Estimation of the parameters of this distribution was carried out with the E-M algorithm (4).

The posterior means of the exposure assessments were computed with the estimated distribution of true measurements, and these were used to find the posterior mean cumulative exposures, which were used to estimate β from model e using maximum likelihood on the set of lung cancer deaths and their smoking matched referents. The effect on the estimate was to increase it from 0.00131 (0.0006) to 0.00161 (0.0007), a result suggesting that this measurement error had biased the estimate downwards by about 20%. This bias is small in comparison with the other uncertainties involved in this type of

study. The effect of the added uncertainty in expressing exposure in terms of fiber concentrations is under investigation.

To check whether the approximation made in using only the posterior means was seriously affecting the estimate of β , a full maximum likelihood estimate was computed. This estimate was 0.00168.

It is not suggested that the estimates of the exposure-response relationship presented are definitive. For example, the confounding effect of smoking, treated otherwise, would give different estimates. Our purpose was solely to show the effect of approximations in exposure assessments.

Acknowledgments

This study was supported by the Medical Research Council, from which one of us (BG Armstrong) received a research studentship, and by a grant from the Quebec Asbestos Mining Association.

References

1. Breslow NE, Day NE. Statistical methods in cancer research. Volume 1. International Agency for Research on Cancer, Lyon 1980.
2. Bross I. Missclassification errors in 2×2 tables. *Biometrics* 10 (1954) 478–486.
3. Cochran WG. Errors of measurement in statistics. *Technometrics* 10 (1968) 637–666.
4. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the E M algorithm. *J r stat soc B* 39 (1977) 1–38.
5. Liddell FDK, McDonald JC, Thomas DC. Methods of cohort analysis: Appraisal by application to asbestos mining. *J r stat soc A* 140 (1977) 469–491.
6. Maritz JS. Empirical Bayes methods. Methuen, London 1970.
7. McDonald JC, Liddell FDK, Gibbs GW, Eyssen GE, McDonald AD. Dust exposure and mortality in chrysotile mining 1910–1975. *Br j ind med* 37 (1980) 11–24.
8. Moran PAP. Estimating structural and functional relationships. *J multivar anal* 1 (1971) 232–255.
9. Oldham PD. The nature of the variability of dust concentrations at the coal face. *Br j ind med* 10 (1953) 227–234.