Sources of Uncertainty in Dose-Response Modeling of Epidemiological Data for Cancer Risk Assessment

LESLIE STAYNER,[a] A. JOHN BAILER, RANDALL SMITH,
STEPHEN GILBERT, FAYE RICE, AND EILEEN KUEMPEL

Education and Information Division, Department of Health and Human Services,
Centers for Disease Control and Prevention,
National Institute for Occupational Safety and Health,
Robert A. Taft Laboratories, 4676 Columbia Parkway, MSC-15,
Cincinnati, Ohio 45226, USA

ABSTRACT: Epidemiologic data is increasingly being used for dose-response analysis in risk assessment. The Environmental Protection Agency (EPA) and other U.S. agencies have expressed a preference for using epidemiologic data rather than toxicologic data when possible. However, there are a number of important sources of uncertainty in using epidemiologic data for this purpose that need to be clearly recognized and, when possible, quantified. This paper presents a critical review of the major sources of uncertainty in the use of epidemiologic data for cancer risk assessment. These may include: (1) study design issues such as potential confounding and other biases, inadequate sample size, and followup, (2) the choice of the data set, (3) specification of the dose-response model, (4) estimation of exposure and dose, and (5) unrecognized variability in susceptibility. Examples from risk assessments for cadmium, asbestos, and diesel exhaust are used to illustrate the potential magnitude of some of these sources of uncertainty. It is shown that the overall uncertainty from these various sources combined may often result in highly uncertain risk estimates from dose-response modeling of epidemiologic data. For this reason, we believe it is best to present a range of possible risk estimates, which, to the extent possible, reflects the variability and uncertainty inherent in the dose-response evaluation of epidemiologic data.

INTRODUCTION

Most regulatory agencies in the United States have expressed a clear preference for using human data, when available, instead of toxicologic data from animal studies, to quantify the risks associated with environmental and occupational carcinogenic hazards. However, most quantitative risk assessments that have been performed to date have been based on dose-response modeling of animal bioassay data. This situation appears to be changing as data from epidemiologic studies become an increasingly important source of information for dose-response modeling

[a]Address for correspondence: 513-533-8365 (voice); 513-533-8224 (fax).
e-mail: lts2@cdc.gov

in the quantitative assessment of human health risks. This change is probably related to the increasing criticism of animal models for predicting human risks;[1] and to improvements in the quality of the available epidemiologic database.

There are several major potential sources of uncertainty that may arise in using epidemiologic data for dose-response assessments. These sources of uncertainty need to be considered, and when possible quantified, in any risk assessment that uses dose-response data from epidemiologic studies. This paper presents a review of the major sources of uncertainty, with examples from risk analyses that have been conducted by the authors.

STUDY DESIGN

The lack of adequate exposure information is probably the most frequently cited limitation of epidemiologic data for risk assessment purposes. Although this certainly is an important factor, which we discuss below, there are several other major limitations on epidemiologic study design that may be as important. Epidemiologic studies, by definition, are observational in nature and, consequently, it is generally impossible to randomize the assignment of exposures. Thus, serious questions often arise about the potential for confounding, selection bias, and other sources of bias. The frequent use of general population rates as the referent group for occupational cohorts often introduces a form of selection bias due to the well known *healthy worker effect*.[2] The impact of this potential bias may be mitigated in many occupational studies, simply by restricting the dose-response analysis to the exposed and non-exposed (if available) subjects within the cohort. However, workers must generally remain healthy to stay employed, a phenomenon that has been termed the healthy worker survivor effect (HWSE),[3] a fact that may exert a strong influence on the dose-response relationship. Steenland *et al.*[4] have shown that a negative bias in exposure-response relationships, attributable to the HWSE, may often be observed in occupational cohort studies.

In addition to potential biases, one has to consider other potential problems arising from limitations in the design of the study. An important question is whether or not the period of observation (follow-up) of the cohort was adequate? Many occupational and environmental cancers have an average latency period of approximately 15 to 20 years, and thus the cohort should be followed for at least this long in order to observe an excess from these health outcomes. One also needs to consider whether the methods for case ascertainment are adequately sensitive. For example, studies based on mortality data may be inadequately sensitive for cancers that are treatable and have long survival times, such as leukemia.

Finally, the size of the study is an extremely important consideration. Sample size estimates for different levels of excess risk are presented in TABLE 1 for a hypothetical study of lung cancer in an occupational cohort. It should be noted that similar results would be obtained for case-control studies nested within a cohort. These sample size estimates are based on a 5% false-positive rate (alpha) and 80% sensitivity (power) for detecting different levels of excess risk. Most cohort mortality studies have at least 1,000 workers and very few have more than 100,000 workers. As this table illustrates, these studies would generally be incapable of detecting an excess

TABLE 1. Sample size estimates for detecting varying levels of excess lung cancer risk in a hypothetical retrospective cohort mortality study

Excess risk	Relative risk (SMR)[a]	Expected deaths[b]	Person years[c]	Number of workers[d]
10^{-2}	1.20	170	217,161	4,343
10^{-3}	1.02	15,605	2.0×10^{7}	399,605
10^{-4}	1.002	1.5×10^{6}	2.0×10^{9}	39.6×10^{6}
10^{-5}	1.0002	1.5×10^{8}	2.0×10^{11}	39.6×10^{10}

[a]Relative risks calculated using a background risk (cumulative probability) of 0.06 for developing lung cancer for males over age 15, based upon the proportion of deaths from lung cancer among U.S. males over age 15 in 1982.

[b]Expected number of deaths calculated assuming 80% power ($1 - \beta$), α level of 0.05 (single tail) and the calculated relative risk.

[c]Person years calculated by dividing the expected number of deaths by the lung cancer rate (7.8×10^{-4}) among males between the ages of 45–54 based upon U.S. mortality rates from 1982 (NCHS 1986) which is approximately the average of the hypothetical population.

[d]Number of workers calculated by assuming each worker contributed 50 person-years to the study.

risk of less than 1 per 1,000 workers. It should also be noted that an excess risk of 1 per 1,000 corresponds to relative risk of 1.02, which most epidemiologists would be highly reluctant to consider meaningful even if it were statistically significant. The U.S. Occupational Safety and Health Administration (OSHA) generally considers a risk of greater than 1 per 1,000 to be significant, and the Environmental Protection Agency (EPA) generally considers a risk of 1 per 1,000,000 to be significant.[5] Thus epidemiologic studies generally have low power for detecting the levels of risk that are of regulatory concern in the United States. Conversely, epidemiologic studies that demonstrate a statistically significant excess are likely to identify risks that are of regulatory concern.

IS THERE A CAUSAL RELATIONSHIP?

Because of the design issues discussed above, it is often difficult, if not impossible, to draw firm conclusions about whether or not a causal association exists between an exposure and disease based on a single or even several epidemiologic studies. Hertz-Picciotto[6] has suggested that epidemiologic data should only be used for extrapolation in risk assessments (that is, for dose-response modeling) if: (1) a moderate to strong positive association exists, (2) strong biases can be ruled out, (3) confounding is well-controlled or limited, and (4) exposures have been well-characterized quantitatively. However, we believe that these criteria, particularly the first, are too restrictive. It may be informative to conduct dose-response analyses using epidemiologic data even if the study is negative or shows only a weak association (i.e., Criterion 1). This type of analysis may at least inform decision makers on what an upper bound estimate or best estimate of risk might be. Similarly a risk analysis based on epidemiologic data may also be informative, even if confounding or other

biases cannot be completely ruled out. For example, such analyses could be used to assess the credibility of other analyses based on animal data or other epidemiologic studies.

CHOICE OF DATA

There is frequently more than one epidemiologic data set that may be used for a dose-response analysis. Not surprisingly, given the issues of study design discussed above, there may be a great deal of heterogeneity in the results from dose-response analyses from different epidemiologic studies. For example, in FIGURE 1 the results from an exposure-response analysis that we performed for chrysotile asbestos and lung cancer, based on a study of textile workers,[7] are contrasted with estimates of risk based on another study of chrysotile miners from Quebec.[8] The slopes from these two studies differ by more than one order of magnitude. It is very difficult for risk assessors and risk managers to deal with this kind of difference, since it is frequently impossible to identify a single study as providing the "best" data for the risk analysis. An example of an approach that has been used is the OSHA final rule for asbestos, where they chose to use the geometric mean of slopes from several studies.[9] An attractive alternative to this problem would be to conduct a meta-analysis or to pool the results from all of the studies available and conduct dose-response anal-

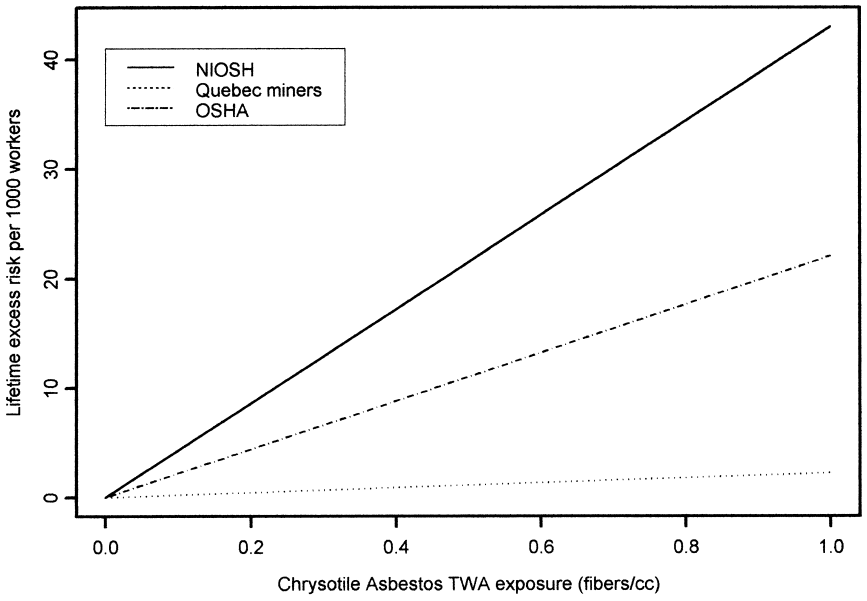

FIGURE 1. Comparison of lung cancer excess risk estimates and chrysotile asbestos exposure based on NIOSH Stayner *et al.* (1997) study of textile workers, McDonald *et al.* (1980) study of Quebec miners and millers, and the OSHA asbestos risk assessment.

yses based on the combined data set. We are currently attempting to perform such a pooled analysis for silica and lung cancer risk in collaboration with scientists at the International Agency for Research on Cancer (IARC).

CHOICE OF THE DOSE-RESPONSE MODEL FORM

Choosing an appropriate dose-response model is a critical step and another major source of uncertainty. In the past, many risk assessments simply assumed a linear relationship between relative risk and cumulative exposure.[10] In part, the justification for this assumption seems to have been based on the multistage theory of carcinogenesis,[11] which suggests that the carcinogenic effects of chemicals would be low-dose linear. However, this model is not truly consistent with an Armitage-Doll model, except when the model has two stages. These models essentially assume the following relationship:

$$RR = 1 + \beta(X)$$

where RR is the relative risk, β is the slope, and X is the cumulative exposure. However, restricting attention to such simple models no longer seems justifiable when modern methods and computing easily permit examination of alternative models with different functional forms.[12] In addition, current theories of carcinogenesis suggest chemicals may act on cell growth and differentiation as well as mutational

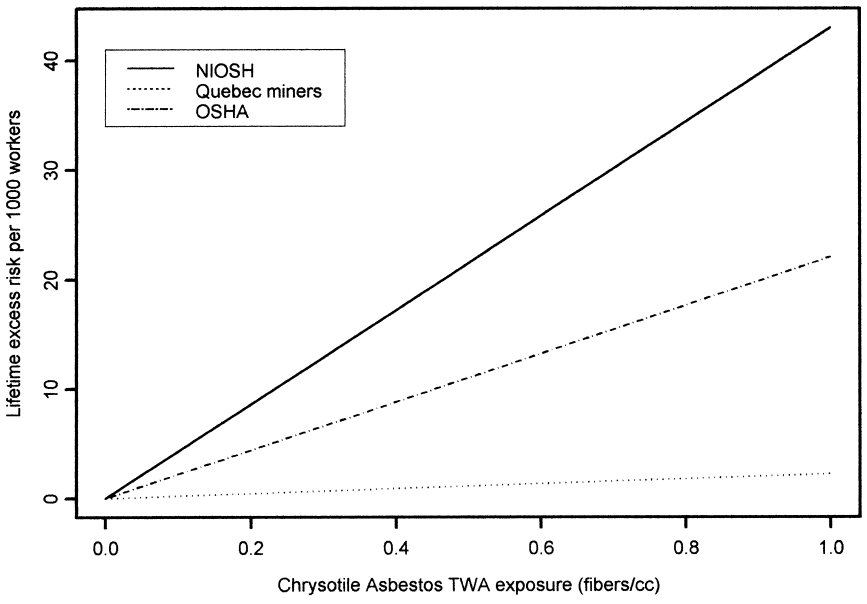

FIGURE 2. Comparison between excess risk estimates derived from various models fitted to lung cancer data from the NIOSH study of cadmium exposed workers. (Adapted from Stayner *et al.* Ref. 12.)

events.[13] Hence a chemical may exert an effect on any number of several available pathways, and the resulting exposure-response relationship may not always be low-dose linear.

The choice of a dose-response model may have a dramatic effect on the resulting estimates of risk, particularly at exposure levels that are well below the levels experienced by subjects of the epidemiologic study. Poisson regression or Cox proportionate hazard models using alternative parametric forms (i.e., log-linear, additive relative risk, and power), as well as biologic models (i.e., multistage or two-stage clonal expansion) may be fitted to the data. In FIGURE 2, estimates of risk are presented for several alternative dose-response models from an analysis performed by Stayner *et al.*[12] of occupational cadmium exposure and lung cancer risk. It can be seen from this figure that the estimates of risk vary by nearly an order of magnitude depending on the dose-response model used.

Splines[14] and other data smoothers offer an attractive new and flexible method for evaluating the shape of the dose-response by making few, if any, parametric assumptions. This method is illustrated in FIGURE 3 from an analysis of chrysotile asbestos and lung cancer risk that we performed.[7] The restricted cubic spline model yielded a very similar fit to the data as that obtained from the additive relative rate model. This result added confidence to our choice of the additive relative rate model as the best parametric form for our risk analysis. Alternatively, one could use the spline model itself for the risk analysis, which in this example would have yielded essen-

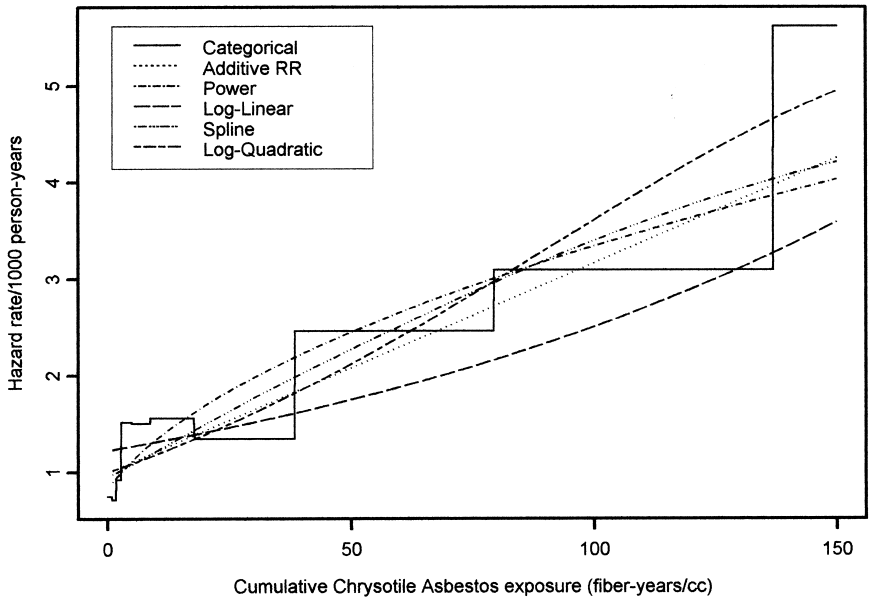

FIGURE 3. Lung cancer mortality rates as a function of cumulative chrysotile asbestos exposure predicted by alternative models for white males, age 50 in 1940–1969. (Adapted from: Stayner *et al.* Ref. 7.)

tially the same result. However, there could be instances where a spline or other smoothers might yield dose-response models with questionable fitted values, since they can be so flexible that they are sensitive to local random variations leading to over fitting the data.

Finally, a critical concern in a risk analysis is whether or not there is a threshold below which the exposure has no effect on the disease risk. The assumption that there is no threshold for carcinogens is increasingly being questioned. Some epidemiologists have argued that there is a threshold at cut points in their data where they no longer see a significant excess risk. This is clearly inappropriate, since there will generally be a level of exposure in an epidemiologic study where one no longer sees an excess risk simply due to a lack of sufficient sample size and associated study power limitations. Furthermore, if exposures are categorized, it may be easily shown that the choice of cut-points (which is generally arbitrary in epidemiologic studies) can influence the determination of the no-adverse-effect level.[15]

Ulm[16] has suggested a formal statistical method for estimating a threshold parameter from epidemiologic data that has continuous exposure information. This is a useful method, although what it yields should probably not be called a "threshold", but rather a point in the data below which there is no evidence of an excess risk. This is because the parameter estimate is influenced by the study design (for example, sample sizes and exposure data) and should not thus be viewed as a true biologic threshold. We applied this method in our analysis of lung cancer risk and exposure to chrysotile asbestos,[7] and we found that the maximum likelihood estimate for this parameter was zero. This may often be the case with epidemiologic data, or the confidence intervals on this parameter may be extremely broad. Even if a true threshold exists, the reality is that epidemiologic studies will seldom be able to detect it. Thus, unless there is other biologic evidence for a threshold, it would be ill-advised to use a threshold model for human health risk assessments. Furthermore, even if there were evidence of a threshold, a single number would not apply to all individuals, and it is more plausible to think in terms of a distribution of thresholds among the population.

ERRORS IN EXPOSURE ESTIMATES

The fact that epidemiologic studies of cancer require information on exposures from 20 or more years prior to the end of the study makes the reconstruction of exposures an extremely difficult exercise, one that is generally fraught with potential errors. Despite this well-recognized fact, most epidemiologic studies and risk assessments based on these studies, treat the exposures as if they were known and without error. It is also commonly assumed in epidemiology that errors in exposure estimates that are non-differential with respect to disease will lead to an underestimate of the true dose-response slope. However, this is not always the case, and errors in exposure may in fact either inflate or deflate the slope of the dose-response relationship depending on the structure of the error.[17]

We are currently working on an analysis of the effects of errors in exposure estimates on the dose-response relationships in a study of railroad workers exposed to diesel exhaust particles.[18] This study has several potential sources of error that may

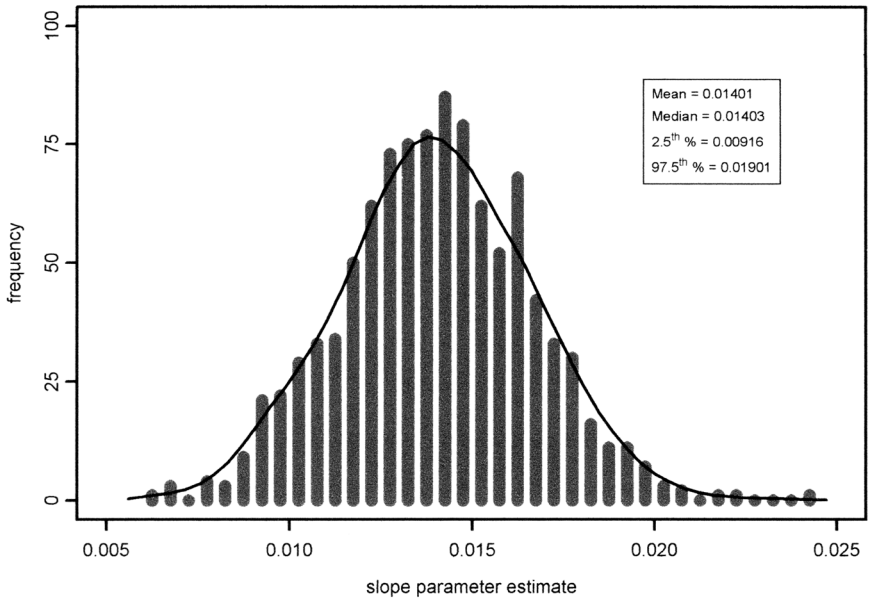

FIGURE 4. Histogram of parameter estimates from 1,000 Monte Carlo simulations of duration of diesel exposure using a linear relative risk model based on data from Garshick *et al.* (1986).

have contributed to a distortion of the dose-response relationship, including errors that effect estimation of both the duration and intensity of exposure. We are exploring the use of Monte Carlo methods to evaluate the potential uncertainty introduced by these exposure estimation errors on quantifying the dose-response relationship. Preliminary results from the Monte Carlo simulation of errors affecting the estimates of duration of exposure are illustrated in FIGURE 4. This figure shows that the slope of the linear relative rate regression model varied by a factor of approximately four (minimum versus maximum) with 95% of the results lying within a factor of two. This range reflects only a part of the overall uncertainty, which is likely to increase as we consider the other sources of uncertainty in the exposure estimates used in this analysis.

ERRORS IN USING EXPOSURE RATHER THAN DOSE

At best, epidemiologic studies of cancer risk rely on estimates of external exposure from personal breathing zone samples and work history information for estimating exposure. The delivered dose, or the actual dose that reaches the target tissue, is rarely available in epidemiologic studies of chronic diseases like cancer. If exposure is proportional to dose, then external exposure would be a reasonable surrogate for

tissue dose, which would differ by some constant factor. However, if exposure is not proportional to dose (for example, when saturation occurs in capacity-limited processes such as uptake, metabolism, or clearance), then external exposure measures would not necessarily represent dose over the entire distribution of exposures. Similarly, if there are systematic differences among individuals in the exposure-dose relationship (for example, genetic polymorphisms resulting in metabolic differences, such as fast or slow acetylators; or pre-existing conditions such as bronchitis, which can alter deposition and clearance in the lungs), then exposure might not be a good representation of dose.

Dose information may be available from autopsy or clinical studies, particularly if the substance is biopersistent. For example, a strong association has been observed in most case-control studies of mesothelioma and dose (lung burdens) of amphibole asbestos, but not of chrysotile asbestos. This may be explained by the fact that chrysotile asbestos has a relatively short half-life in the lung; whereas, amphiboles have a long half-life.[19]

Dosimetric models have been developed for a relatively few occupational and environmental epidemiologic studies.[20] Kuempel[21] recently developed a human dosimetric lung model of the long-term retention of particulates, using autopsy data of U.S. coal miners, information on job-specific duration and intensity of exposure to respirable coal mine dust, and other human data on breathing rates, particle deposition in the lungs, and initial clearance rates. She found that measured lung dust burden was a stronger predictor of the probability for developing pulmonary fibrosis than was cumulative exposure, and that the model-predicted lung burdens also showed statistically significant dose-response, with similar coefficients to those for measured lung dust burdens. Furthermore, she found a different pattern of exposure-dose in humans than that observed in animal studies with chronic exposures to particles. This illustrates the potential usefulness of dosimetric models in risk assessment, which can represent biologic processes that affect dose.

Clearly our inability to use biologic markers for dose and our reliance on using measures of external exposure introduces uncertainty into the use of epidemiologic data in the risk assessment process. It is difficult at this time to judge the extent of this uncertainty, but it may be large in some cases, particularly when individual characteristics are known to modify the absorption, metabolism and delivery of the exposure to the target tissue.

HUMAN VARIABILITY IN SUSCEPTIBILITY

The effects of human variability in susceptibility to exposure on the results from risk analyses from epidemiologic studies have largely been neglected to date. We generally fit our models to epidemiologic data by assuming all of the individuals in the study are from the same population, and then extrapolate our findings to other populations who may have quite different characteristics that effect susceptibility. The reason we have done so is simply because of our near total ignorance of what these factors are, and how they are distributed in our study population. We are becoming increasingly aware of how bad this assumption is, and of the existence of subpopulations in our studies with genetic polymorphisms that influence their risk.

In fact, it may often be the case that the individuals at the greatest risk are those that have some combination of different genetic polymorphisms. For example, in a recent case-control study of lung cancer, Hirvonen[22] evaluated interactions between asbestos exposure, GSTM1 genotype, and N-acetyltransferase slow acetylator genotype (NAT-2). Being either $GSTM1^{null}$ or a slow acetylator ($NAT\text{-}2^{*slow}$) was associated with an approximately twofold increased risk of lung cancer. Having both *at-risk* genotypes ($GSTM1^{null}$ and $NAT\text{-}2^{*slow}$) was associated with an approximately fourfold increase in risk. Having both *at-risk* genotypes and being highly exposed to asbestos was associated with an approximately eightfold increase in risk. Hence the results from this study suggest a multiplicative relationship between *GSTM1, NAT-2* and high asbestos exposure. This example illustrates how large the differences in risk may be for subpopulations in our studies, and in the populations for which we are trying to estimate risk. Our failure to recognize these difference may lead to large errors in our risk estimates particularly for certain members of the population.

CONCLUSION

In this paper we have attempted to briefly discuss and illustrate some of the major sources of uncertainty in using epidemiologic data for dose-response analyses. Several of these sources have the potential to result in relatively large errors in the estimation of risk. Our cadmium example illustrates that just varying the statistical model may result in risk estimates that span an order of magnitude. Other sources of uncertainty reviewed in this paper may also easily result in errors in predicted risks that are as large. The overall uncertainty from these various sources combined may often result in risk estimates from dose-response modeling of epidemiologic data that are highly uncertain. For this reason, we believe that it is best to present a range of possible risk estimates, which, to the extent possible, reflects the variability and uncertainty inherent in the dose-response evaluation of epidemiologic data.

REFERENCES

1. AMES, B.N. & L.S. GOLD. 1990. Chemical carcinogenesis: too many rodent carcinogens. Proc. Natl. Acad. Sci. **87:** 7772–7776.
2. MCMICHAEL, A.J. 1976. Standardized mortality ratios and the "Healthy Worker Effect": scratching beneath the surface. J. Occupational Medicine **18**(3): 165–168.
3. ROBBINS, J. 1990. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. J. Chronic Dis. **40**(Suppl. 2): 1091–1098.
4. STEENLAND, K., J. DEDDENS, A. SALVAN & L. STAYNER. 1996. Negative bias in exposure-response trends in occupational studies: Modeling the healthy workers survivor effect. Am. J. Epi. **143**(2): 202–210.
5. RODERICKS, J.V., S.M. BRETT & G.C. WRENN. 1987. Significant risk decisions in federal regulatory agencies. Regulatory Toxicol. and Pharmacol. **7:** 307–320.
6. HERTZ-PICCIOTO, I.H. 1995. Epidemiology and quantitative risk assessment: a bridge from science to policy. Am. J. Public Health **85:** 484–491.
7. STAYNER, L.T., R. SMITH, J. BAILER, S. GILBERT, K. STEENLAND, J. DEMENT, D. BROWN & R. LEMEN. 1997. Exposure-response analysis of respiratory disease risk associated with occupational exposure to chrysotile asbestos. Occupational Environmental Medicine **54:** 646–652.

8. MCDONALD, J.C., F.D.K. LIDDELL, G.W. GIBBS, G.E. EYSSEN & A.D. MCDONALD. 1980. Dust exposure and mortality in chrysotile mining, 1910-75. Br. J. Ind. Med. **37:** 11–24.
9. OSHA. 1986. Occupational exposure to asbestos, tremolite, anthophylite and actinolite. Federal Register **51:** 22612–22747.
10. SMITH, A.H. 1988. Epidemiologic input to environmental risk assessment. Arch. Environ. Health **43:** 124–127.
11. ARMITAGE, P. & R. DOLL. 1954. The age distribution of cancer and a multistage theory of carcinogenesis. Br. J. Cancer. **VIII**(1): 1–12.
12. STAYNER, L.T., R. SMITH, A. J. BAILER, G. E. LUEBECK & S.H. MOOLGAVKAR. 1995. Modeling epidemiologic studies of occupational cohorts for the quantitative assessment of carcinogenic hazards. Am. J. Ind. Med. **27:** 155–170.
13. MOOLGAVKAR, S.H. & G.E. LUEBECK. 1990. Two-event model for carcinogenesis: Biological, mathematical, and statistical considerations. Risk Analysis **10:** 323–341.
14. HASTIE, T.J. & R.J. TIBSHIRANI. 1990. Generalized Additive Models. Chapman & Hall, London.
15. BAILER, A.J., L.T. STAYNER, R.J. SMITH, E.D. KUEMPEL & M.M. PRINCE. 1997. Estimating benchmark concentrations and other noncancer endpoints in epidemiology studies. Risk Analysis **17:** 771–780.
16. ULM, K.W. 1990. Threshold models in occupational epidemiology. Mathematical Computer Modeling **14:** 649–652.
17. Wacholder, S. 1995. When measurement errors correlate with truth: Surprising effects of nondifferential misclassification. Epidemiology **6**(2): 157–161.
18. GARSHICK, E., M.B. SCHENKER, A. MUNOZ, M. SEGAL, T.J. SMITH, S.R. WOSKIE, S.K. HAMMOND & F.E. SPEIZER. 1988. A retrospective cohort study of lung cancer and diesel exhaust exposure in railroad workers. Am. Rev. Resp. Dis. **137:** 820–825.
19. STAYNER, L., D.A. DANKOVIC & R.A. LEMEN. 1996. Occupational exposure to chrysotile asbestos and cancer risk. A review of the "amphibole hypothesis". Am. J. Public Health **86**(2): 179–186.
20. KREIBEL, D. 1994. The dosimetric model in occupational and environmental epidemiology. Occ. Hyg. **1:** 55–68.
21. KUEMPEL, E. 1997. Development of a biomathematical lung model to describe respirable particle retention and to investigate exposure, dose and disease in U.S. coal miners. Dissertation, University of Cincinnati, Cincinnati, Ohio.
22. HIRVONEN, A. 1997. Combinations of susceptible genotypes and individual responses to toxicants. Environmental Health Perspectives **105:** 755–758.