Meta-analysis and Risk Assessment

Catherine Wright
Peggy Lopipero
Allan Smith

The most important questions about the relationship between environmental agents and human disease are often the most difficult to answer. Does the weight of the evidence indicate that exposure causes disease? If so, what is the quantitative dose-response relationship? What does that relationship imply for regulation? Risk assessment is usually defined as the process by which these questions are answered. Within the terminology of risk assessment, the first question is often called hazard identification, the second concerns dose-response assessment, and the third question combines estimates of the prevalence of exposure with risk quantification. Risk assessment can be done with animal or human data.

Meta-analysis is a tool often used in risk assessment to combine results across studies, usually human studies, with the goal of estimating measures of association with improved precision. It is commonly used at the hazard identification stage to combine a number of studies of exposed versus nonexposed populations and provide a quantitative summary of results. It is sometimes also used to assess dose-response relationships when there are several studies with dose-response results and one wishes to combine them.

Meta-analysis is increasingly replacing the traditional review of the literature because of its ability to provide a quantitative estimate with corresponding confidence intervals. It is done by calculating a weighted average of results across studies, with the weights usually being the inverse of the variance of the result for each study (i.e., larger studies with more precise results are given greater weight).

A problem with meta-analysis is its tendency to combine apples and oranges; well-designed studies may be combined with poorly designed ones. One way to confront this problem is to conduct analyses of subsets of studies that are deemed to be better designed versus those less well designed. Results may help decide whether an exposure truly causes disease. For example, if better-designed studies with good control over confounding show no exposure effect but less well designed studies do show such an effect, one might conclude that confounding was responsible for biased results in the latter studies and that exposure did not cause disease.

A second problem in meta-analysis is publication bias, meaning that "negative" studies tend not to be published and thus are unavailable for the meta-analysis. This produces results that may be biased toward finding an effect. There is no ideal solution to this problem beyond searching diligently for all studies on a given topic, including if possible any that are lying on some investigator’s back shelf.

Once a hazard has been identified, typically by a meta-analysis, risk assessment has the more ambitious goal of quantifying potential health risks in relation to exposure. Typically this involves determining a dose-response relationship, and then using the resulting excess risk associated with each unit of exposure to estimate an acceptable level of exposure over a lifetime.

By definition, the process of risk assessment is a speculative exercise, laden with assumptions. The underlying question of the quantitative impact of exposure on disease is truly the bottom line for all environmental health research. Instead of hedging with vague terms of "small risk" or "likely to be substantial," risk assessment seeks the best possible quantitative estimate of the risk. Debate over risk assessment is predictable, given the direct application to regulatory policy and untidiness inherent in the generation of quantitative results. Informal or implicit risk assessment is part of everyday life at the individual and societal level. The unique feature of formal risk assessment is its explicitness in delineating each step in the process and the assumptions involved for each step.

One of the important products of risk assessment, in addition to the quantitative results, is the identification of critical gaps in our knowledge that prevent the assessment from being more definitive. Among the lists of assumptions and data sources, there are often just a few key issues that are the dominant sources of uncertainty. Most commonly, in risk assessments based on animal data, the limitation is in extrapolation across species and from high to low doses. In risk assessments based on epidemiologic data, estimates of the level of exposure for study subjects—key to dose-response analyses—are often subject to substantial uncertainty. Data on confounders may also be minimal. It is not always appreciated that the proper comparison is not between the clarity of toxicologic...
studies compared to epidemiologic studies, which must favor the toxicologic research because of tight experimental control. The real question is the information value for human risk assessment, and perfect toxicology may be roughly equivalent to adequate epidemiology, with high-quality epidemiology superior to all other approaches.

The methods of risk assessment have evolved principally to address cancer, and extrapolation from animals to humans is most advanced for addressing this set of health end points. The combination of extensive experimental results, a substantial body of epidemiologic findings, and some general ideas of mechanisms makes risk assessment for cancer relatively advanced, at least compared to the kind of assessment that is possible for such outcomes as reproductive health or neurological disease. There may be some circular reasoning, that regulation is most often based on carcinogenicity, and the tools for risk assessment, which serves as the basis for regulation, are most fully developed to address carcinogenicity. However, the general principles of risk assessment are applicable to all health outcomes, and creative approaches are needed to apply those principles to a more comprehensive array of health considerations.

Environmental risk assessment is the quantification of potential adverse health effects of human exposure to environmental hazards. To estimate health risks at low environmental exposure levels often requires extrapolating from studies with high exposures to chemical or physical agents. High-exposure studies include epidemiologic studies of occupationally exposed cohorts, epidemiologic studies involving high environmental exposures in the community, and experimental animal studies. Results of risk assessments are used by risk managers to set regulatory policies regarding acceptable levels of chemical contaminants in the environment.

Risk assessments contain some or all of the following four steps: hazard identification, dose-response assessment, exposure assessment, and risk quantification (EPA, 1989). Hazard identification involves determining whether an agent can cause an adverse health outcome, such as cancer or other disease. Dose-response assessment involves determining the relationship between the magnitude of exposure and the probability of occurrence of the health effects in question. Exposure assessment is the evaluation of the extent of public exposure to the chemical or physical agent; it may include environmental or occupational exposure measurement, emission or effluent quantification, modeling of environmental transport and fate, identification of exposure routes, identification of exposed populations, and estimation of short-term and long-term exposure levels. Risk quantification requires calculation of the magnitude of human risk at various exposure levels, and often includes an analysis of uncertainty inherent in the risk estimates.

Risk assessments may be based on experimental animal data or epidemiologic data. Although epidemiologic data are considered the most convincing evidence of human risk in the regulatory setting, risk assessments are often performed with animal data because adequate human data are often lacking. There are a number of advantages and disadvantages in conducting risk assessment based on either type of data.

The advantages of using long-term studies in animals for risk assessment are that exposure levels and conditions are known, and the toxicity and carcinogenicity of individual chemicals can be clearly identified. However, animal experiments are conducted at exposure levels far in excess of those anticipated for humans, and they are often conducted using a route of exposure that may not be relevant to human exposure scenarios. Animal experiments also do not reflect human exposure circumstances, because animal tests are typically conducted with single chemicals whereas humans experience multiple exposures during their entire lives (Huff and Hoel, 1992). The relevance of animal experiments to humans is also questionable because of biological differences between humans and laboratory animals.

Little is known about the true correlation between carcinogenic effects in animals and those in humans. Allen et al. (1988) argue that there is a relatively good correlation between known human carcinogens and animal carcinogenicity. This conclusion is based on a limited number of chemicals, however. Some work has been done looking at interspecies correlations between rats and mice (Piegorsch et al., 1992; Haseman and Lockhart, 1993). If mice and rats are similar in regard to carcinogenesis, this provides some evidence in favor of interspecies extrapolations. Haseman and Lockhart (1998) examined a database of 379 long-term carcinogenicity studies in rats and mice to evaluate sex and species correlations in site-specific carcinogenesis response. Within a species, target sites showed only 65% agreement between males and females. The overall concordance in carcinogenic response between rats and mice was 74% when all target sites were considered collectively. However, the correlation between rats and mice in site-specific carcinogenic response was only 37%. In fact, no significant interspecies correlations were observed for lung tumors, which are a common concern in environmental epidemiology and health risk assessment.

A number of articles have also been published on the correlations between carcinogenic potencies in rats and mice. Crouch and Wilson (1979) demonstrated a strong correlation between carcinogenic potencies in rats and mice, supporting the extrapolation from mouse to humans.
Although there are a number of advantages in doing meta-analysis in risk assessment, the approach can result in inaccurate conclusions if not properly conducted.

This chapter focuses on the hazard identification and dose-response assessment phases of health risk assessment. An example of the use of meta-analysis in hazard identification will be presented as well as a dose-response assessment utilizing data from occupational epidemiologic studies.

**Hazard Identification**

Hazard identification involves determining if adverse health effects are caused by a particular exposure. Its purpose in environmental epidemiology and health risk assessment is to conclude whether a chemical or physical agent is a human carcinogen or causes other types of adverse health effects. Hazard identification typically includes a review of the physical-chemical properties of the chemical and routes and patterns of exposure; structure-activity relationships; metabolic and pharmacokinetic properties; toxicologic effects other than cancer; short-term tests; long-term animal studies; and human studies. Evidence of possible carcinogenicity in humans comes primarily from long-term animal tests and epidemiologic studies.

**Using animal data for hazard identification**

In the absence of adequate human evidence, evidence from animal experiments is used to identify chemical hazards. Criteria for the technical adequacy of animal carcinogenicity studies have been published by the NTP (National Toxicology Program, 1984) and the EPA (Environmental Protection Agency 1983a,b,c). The weight of evidence that an agent is potentially carcinogenic for humans increases (1) with the increase in number of tissue sites affected by the agent; (2) with the increase in number of animal species, strains, sexes, and number of experiments and doses showing a carcinogenic response; (3) with the occurrence of clear-cut dose-response relationships (malignant and benign tumors combined) in treated compared with control groups; (4) when there is a dose-related shortening of the time-to-tumor occurrence or time to death with tumor; and (5) when there is a dose-related increase in the proportion of tumors that are malignant (EPA, 1989).

A number of factors need to be considered when reviewing the weight of evidence in long-term animal studies. For example, long-term animal
studies are typically conducted at or near the maximum tolerated dose (MTD) level to ensure adequate statistical power for the detection of carcinogenic activity. However, the applicability of these bioassay results to human conditions has been questioned because qualitatively different biologic responses may occur at very high exposure levels. Critics have argued that the testing of animals at the MTD and at one-half the MTD causes inflammation and cell proliferation that does not occur at low exposures and that elevated rates of cell mitosis increase the opportunities for spontaneous mutations, thereby contributing to carcinogenic development (Ames and Gold, 1990a, b, 1991). Advocates of the testing protocol state that: (1) toxicity, although frequently observed at the MTD, is usually not observed at one-half MTD even though increased tumor incidence usually is and (2) 90% of the carcinogens identified by the NTP induced tumors in organs that showed no evidence of cellular toxicity (Infante, 1991; Perera 1990; Huff and Haseman, 1991; Rall, 1991; Cogliano et al., 1991).

Tumor data from sites with high spontaneous background rates also require special consideration. For example, there are widely diverging scientific views about the validity of mouse liver tumors as an indication of potential carcinogenicity in humans when such tumors occur in strains with high spontaneous background incidence and when they constitute the only tumor response to an agent (EPA, 1989).

Using epidemiologic studies for hazard identification

The decision as to whether an epidemiologic study is appropriate as the basis for hazard identification involves judgments about both the qualitative and quantitative nature of the data. Qualitative factors that should be considered include the appropriate selection of the exposed and comparison groups, the reliability of exposure ascertainment, the completeness of follow-up, and the potential biases. If a study meets the minimum criteria for acceptance on qualitative grounds, its appropriateness for quantitative risk assessment depends on its ability to yield reasonably reliable information on doses, time and duration of exposure, and magnitude of response. Unfortunately, the magnitude of past exposure often has to be estimated from fragmentary information. The reliability of quantitative information on the magnitude of response depends on the statistical power of the study, on the completeness and reliability of the information on disease incidence, and on the ability to control for confounding factors, such as cigarette smoking. These qualitative and quantitative criteria are more likely to be met for occupational epidemiologic studies than for environmental epidemiologic studies. For these reasons, quantitative risk assessment using human data is largely based on occupational studies.

In hazard identification, epidemiologic data are used to infer causal associations. The key factors to consider in determining causality from epidemiologic studies are chance, bias, consistency, strength, dose response, temporality, and plausibility.

**Chance:** How likely is it that the findings in all available studies are due to chance?

**Bias:** Potential sources of bias should be identified, including selection bias, information bias, and confounding bias. Then, the direction and magnitude of the bias should be determined to assess their effects on study findings. Many potential biases in epidemiological studies may be quite small and of known direction, so it is not appropriate to dismiss a study merely because of the possibility of a particular bias.

**Consistency:** Studies should demonstrate similar associations that persist despite differing circumstances. This does not mean that studies should produce identical findings, because fluctuations may be caused by chance. In addition, fluctuations are expected because of differences in the durations and intensities of exposure, differences in length of follow-up, and variation in susceptibility between populations.

**Strength:** Strength of association is an important criterion, because the greater the estimate of risk and the more precise (narrow confidence limits), the more credible the causal association.

**Dose response:** Dose response is observed if the increase in the measure of effect is positively correlated with an increase in the exposure or estimated dose. A strong dose-response relationship across several categories of exposure supports a causal relationship. However, the absence of a dose-response relationship should not be construed by itself as evidence of a lack of a causal relationship.

**Temporality:** A temporal relationship must be present if causality is to be considered. This means that the disease occurs within a biologically reasonable time frame after the exposure to account for the specific health effects. Some cancers, for example, have latency periods ranging from 20 to 40 or more years from initial exposure.

**Biological plausibility:** The association should make sense in terms of biological knowledge. Information from toxicology, pharmacokinetics, genotoxicity and in vitro studies should be considered.

All the above factors should be considered in making causal inference from epidemiologic studies. Causal inference can usually be made only
when each of these criteria is met by at least some of the studies. However, inferring causation may be reasonable even in the absence of dose-response information. There should always be a degree of consistency between studies, but it is quite common to see the conclusion made that there is a lack of consistency when, in fact, there is a good explanation for apparent inconsistency. In particular, one should not expect consistency in relative risk estimates when there is wide variation in intensity and duration of exposure between studies.

Dose-Response Assessment

Dose-response assessment is the process of characterizing the relationship between the exposure to an agent and the incidence of an adverse health effect in exposed populations. In quantitative cancer risk assessment, the dose-response relationship is expressed in terms of a linear slope (called a potency slope), which is used in the risk characterization phase to calculate the probability or risk of cancer associated with a given exposure level. One of the primary differences in assessing the dose response for cancer vs. noncancer health outcomes is that carcinogens are commonly assumed to have no threshold whereas noncancer effects are assumed to have a threshold. These assumptions are made by regulatory bodies performing a health risk assessment unless evidence suggests otherwise. The focus of this section is dose-response assessment for carcinogenic effects. Methods for noncancer dose-response assessment have been described by the EPA (1999).

Dose-response assessment using animal data

In the absence of appropriate human studies, data from an animal species that responds most like humans can be used in the dose-response assessment. The EPA (1989) has established guidelines for selecting the appropriate set of data for the assessment, assuming several studies are available to choose from. First, the tumor incidence data are separated according to organ site and tumor type. Second, all biologically and statistically acceptable data sets are presented. Third, the range of the risk estimates is presented with consideration of the biological relevance and appropriateness of route of exposure. Finally, because it is possible that human sensitivity is as high as the most sensitive responding animal species, the biologically acceptable data set from long-term animal studies showing the greatest sensitivity (i.e., the highest potency) is generally given the greatest emphasis.

Low-dose estimates derived from experimental animal data extrapolated to humans are complicated by a variety of factors that differ among species and potentially affect the response to carcinogens. Included among these factors are differences between humans and experimental test animals with respect to life span, body size, genetic variability, population homogeneity, existence of concurrent disease, pharmacokinetic effects such as metabolism and excretion patterns, and the exposure regimen. Extrapolations may also be necessary for route of exposure when the exposure route in the animal study selected for dose-response assessment differs from the route of exposure expected in humans (EPA, 1989).

Equivalent doses between species (animal-to-human dose conversions) may be expressed as mg/kg/body weight/day, parts per million (ppm) in diet or water, mg/m² surface area per day, or mg/kg/body weight/lifetime. The equivalent dose generally used by the EPA is mg/m² surface area/day (EPA, 1987). The reason for selecting the surface area conversion is that certain pharmacologic effects, particularly metabolic rate, commonly vary according to surface area (EPA, 1987). The comparison of an effect between species is proportional to dose/body surface, and body surface is proportional to an animal's weight to the two-thirds power. Thus, mg/weight⁵/₃/day is considered an equivalent dose between mammalian species in the absence of better information on pharmacokinetic differences between animals and humans (EPA, 1997). For example, if a rat is exposed to 100 mg/day, then an equivalent dose for a human for the same exposure would be:

\[
\frac{100}{0.35^\frac{5}{3}} \text{ kg} = X / 70^\frac{5}{3} \text{ kg}
\]

Where:

\[
X = \text{equivalent human dose (mg / day)}
\]

\[
0.35 = \text{weight of rat (kg)}
\]

\[
70 = \text{weight of human (kg)}
\]

Solving for X gives:

\[
\frac{100}{0.49} = X / 17
\]

\[
X = 5000 \text{ mg / day}
\]

Because risks at low exposure levels cannot be directly measured by high-dose animal experiments, mathematical models are used that specify the form of the dose-response relationship at low doses. The choice of model can have a large impact on the final risk estimate, particularly if the human exposures of interest are as much as 100 or 1000 times lower than the doses used in the animal experiments. The discrepancies resulting from the use of different models can be 1000-fold or greater.
Biologically based models, particularly models based on the Armitage-Doll (1954) multistage theory of carcinogenesis, have generally been used for producing low-dose risk estimates from animal bioassay data. The multistage theory asserts that in order for a cell to become cancerous it must progress through a series of ordered, independent, and irreversible stages. This model is approximately linear in the low dose region of the dose-response curve and is therefore thought to be relatively conservative (EPA, 1993). The version of the linearized multistage model most commonly employed by the EPA was developed by Crump et al. (1977) and is expressed as follows:

\[ P(d) = 1 - \exp[-(q_0 + q_1d + q_2d^2 + \cdots + q_md^m)], \text{ } k \geq 1 \]

where:

- \( P(d) \) = the probability of cancer at dose \( d \);
- \( k \) = the number of stages, or \( k \) may also be assumed to be equal to the number of dose levels minus one;
- \( q_k \) = coefficients fitted to the data, and;
- \( d^m \) = the applied dose raised to the \( k \)th power.

The upper 95% confidence interval (CI) estimate for \( q_i \) is then used in calculating comparable human risk. More recently, another biologically based model that takes into account effects on cell proliferation, called the two-stage clonal expansion model, has been increasingly used (Moolgavkar and Knudsen, 1981). These and other dose-response models have been described in more detail by the EPA (1993).

**Dose-response assessment based on epidemiologic studies**

Dose-response estimates based on adequate positive epidemiologic data are preferred over estimates based on animal data. The criteria for selecting an epidemiologic study for dose-response assessment includes (1) consistency of findings with other studies, (2) quality of exposure data for relevant period, (3) statistical precision of the risk estimates, (4) dose-response data, (5) data concerning major confounding factors, and (6) adequacy of follow-up in cohort studies.

The use of epidemiologic data in risk assessment should not be based solely on comparison of the quality of available human exposure data to that in animal studies (Smith, 1988). Even if the exposure data are poor, the fact that no species extrapolation is necessary makes the use of human data preferable over animal data. There are, in fact, surrogate measures for exposure that can be used, such as duration of exposure where data on mean exposure levels are available (Shore et al., 1992; Enterline, 1987).

Epidemiologic studies used in risk assessment generally involve high exposures to the agents of concern, thus requiring extrapolation to risks at low exposure levels. Linear dose-response assumptions for low doses have been used extensively for cancer risk assessment from epidemiologic data.

One simple method for modeling dose response when no dose-specific data are available is to plot exposure versus a relative risk estimate such as the standardized mortality ratio (SMR) by assigning a single average exposure level for the entire cohort and drawing a line from the observed SMR to the origin (SMR = 1) (Smith, 1988). This approach is essentially equal to fitting the model \( \text{SMR} = 1 + x\beta \), where \( x = \text{exposure} \) and \( \beta = \text{the change in the SMR per unit of exposure} \) (i.e., the slope).

When dose-specific data are available, such as when SMRs are presented for different levels of cumulative exposure, a simple model is to fit the data with weighted least squares regression, forcing the line to go through an SMR of 1 for zero exposure. This model may also be represented as \( \text{SMR} = 1 + x\beta \), where \( \beta \) is obtained from weighted least squares regression.

It is also possible to apply large numbers of different statistical models each with different assumptions. Normally, however, very few data points are available, and because extrapolations have to be made far below the observed data points, different models can obviously produce markedly different results. Some investigators propose using a variety of different models and giving a range of results (Stayner et al. 1994). However, this approach creates serious problems in the use of human data for health risk assessment (as it does for use of animal studies). In our view, linear regression of relative risk estimates, such as the SMR, forced through a relative risk of 1 for zero exposure, provides stable reproducible results, and is the preferred method for regulatory purposes unless there is clear nonlinearity in the data.

There are a number of reasons for proposing the relative risk model. First, relative risk estimates are usually given in published studies. This means that risk assessments can be conducted without getting the original study data. Second, relative risk estimates in the published literature have already been adjusted for age. All diseases are strongly related to age, therefore modeling risk estimates by other approaches such as additive risk models (\( \lambda_e = \lambda_0 + x\beta \), where \( \lambda_0 \) = rate at exposure \( x \), and \( \lambda_e \) = rate in unexposed) need to incorporate complex functions of age. Another reason is that when relative risk estimates are plotted against cumulative exposure, the relationship is usually linear or close to it. There are not
usually sufficient data points to reject linearity. It should be noted that apparent nonlinearity at low exposure points in cohort studies can be fitted with statistical models that have a profound impact on risk extrapolations to lower doses. However, the empirical evidence for nonlinearity may be extremely weak. Finally, there are often no good biological reasons for rejecting linearity. For these reasons it would seem preferable to use the linear relative risk model for quantitative risk assessment using epidemiologic data, unless there are good reasons to reject it (i.e., clear evidence of nonlinearity). This does not mean that we criticize the investigation of other models for research purposes. However, for regulatory purposes we believe that this model should be the first choice.

There is one situation in which the relative risk model cannot be used. This occurs when the background rate of the disease in question is extremely low. In these cases, relative risk estimates are very unstable. For example, asbestos is by far the main cause of mesothelioma in adults. The background rates without asbestos exposure are extremely low. SMR estimates are therefore very large, and they are unstable because the expected numbers of cases for a cohort are very small (usually a small fraction of 1). In these settings, additive risk models need to be used, but they will not be discussed here.

An example of dose-response assessment for airborne nickel and lung cancer using published epidemiologic data is presented below. This is an example where one study proved superior to others for dose-response assessment as opposed to the combining of several study findings by metaanalysis.

**Example: Dose-response assessment for airborne nickel and lung cancer using published epidemiologic data**

In the following example of risk assessment, published data on occupational cohorts exposed to airborne nickel were used to determine the risk of lung cancer in nickel-exposed communities. This dose-response assessment involves linear regression on SMRs and cumulative exposure estimates. The ultimate goal of this dose-response assessment is to establish a unit risk estimate (URE) that can be used to estimate the cancer risk to any community exposed to nickel in air. The unit risk estimate is currently defined as an estimate of the increased cancer risk from a lifetime (70-year) exposure to a concentration of one unit exposure (EPA, 1995). The URE for inhalation is expressed as risk per μg/m³ for air contaminants.

Based on the criteria described earlier in this chapter, four published cohort studies of nickel refinery workers were considered candidates for risk assessment purposes (Enterline and Marsh 1982; Magnus et al., 1982; Morgan, 1985; Chovil et al., 1981). Table 3-1 summarizes data from the four studies. The ratio of the excess relative risk point estimate (SMR - 1) divided by the excess relative risk at the upper confidence limit of the SMR gives an "index of precision" by which studies can be compared for use in risk assessment. The West Virginia study (Enterline and Marsh, 1982) was found to be unsuitable for risk estimation because of a low index of precision. The Norwegian (Magnus et al., 1982) and Welsh (Morgan, 1985) studies, although having high indices of precision, were found to be unsuitable because of their lack of exposure data.

The Ontario cohort study was determined to be the most appropriate for cancer risk estimation (Chovil et al., 1981). Exposure measures were available for the whole period of operation of the plant from 1948 to 1963, and the index of precision was high (Table 3-1). A weakness of this study was the incomplete follow-up (75%) which has led some to reject the study from serious consideration (Grandjean et al., 1988). However, bounds can be put on the uncertainty in risk estimation due to incomplete follow-up. The overall strengths of this study led to its choice for lung cancer risk estimation. Although more than one study can be used in a dose-response analysis, in this case only the one study had acceptable data.

The Ontario cohort involved refinery workers at Copper Cliff. Follow-up of this cohort, including the group of sinter plant workers used in this risk assessment, was continued to 1984 (ICNRM, 1990). This plant is unique in that there were measurements relating to airborne nickel made back to 1948 when the plant was opened. Levels of nickel in air escaping from the roof monitors of this plant are available for the period 1948 to 1962 (Warner, 1985).

<table>
<thead>
<tr>
<th>Table 3-1. Studies considered for quantitative risk assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>West VA</td>
</tr>
<tr>
<td>(Enterline and Marsh, 1982)</td>
</tr>
<tr>
<td>Norway</td>
</tr>
<tr>
<td>(Magnus et al., 1982)</td>
</tr>
<tr>
<td>Wales</td>
</tr>
<tr>
<td>(Morgan, 1985)</td>
</tr>
<tr>
<td>Ontario</td>
</tr>
<tr>
<td>(Chovil et al., 1981)</td>
</tr>
</tbody>
</table>

*Ratio of (SMR-1) to upper confidence limit of (SMR-1), where (SMR-1) is an estimate of excess relative risk.

* SIR (standardised incidence ratio)
Table 3-2. Observed and expected numbers of deaths and standardized mortality ratios for the Ontario nickel refinery cohort by cumulative exposure

<table>
<thead>
<tr>
<th>Cumulative Exposure (mg/m³) × Years</th>
<th>Observed Lung Cancer Deaths</th>
<th>Expected Lung Cancer Deaths</th>
<th>Standardized Mortality Ratios (SMR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>0</td>
<td>0.47</td>
<td>0.0</td>
</tr>
<tr>
<td>150</td>
<td>0</td>
<td>0.36</td>
<td>0.0</td>
</tr>
<tr>
<td>300</td>
<td>3</td>
<td>0.54</td>
<td>7.5</td>
</tr>
<tr>
<td>490</td>
<td>4</td>
<td>0.60</td>
<td>8.9</td>
</tr>
<tr>
<td>710</td>
<td>6</td>
<td>0.68</td>
<td>11.7</td>
</tr>
<tr>
<td>940</td>
<td>13</td>
<td>0.76</td>
<td>22.8</td>
</tr>
<tr>
<td>1200</td>
<td>11</td>
<td>0.84</td>
<td>17.5</td>
</tr>
</tbody>
</table>

Dose-Response Calculations for the Ontario Cohort Study

The next step in the risk assessment is to determine the quantitative relationship between the estimated dose of nickel and lung cancer relative risk (i.e., the response) for the Ontario cohort. The SMR data for the Ontario cohort study are presented in Table 3-2. The expected numbers of deaths were calculated by Chovil et al. (1981) from age-specific rates for males in Ontario. The cumulative exposure data in the first column of Table 3-2 were derived by the EPA (1986) from exposure duration data (Chovil et al., 1981) incorporating a level of 200 mg/m³ for work prior to 1952, and 100 mg/m³ for 1952 and thereafter (Warner, 1985). Reexamination of the Warner data led us to the conclusion that the average exposure level was actually around 75 mg/m³ from 1952 on, and about 150 mg/m³ before 1952. To account for this difference we multiplied the cumulative exposure estimates by an adjustment factor of 0.75. These adjusted cumulative exposures are shown in Table 3-2.

One criticism of the Chovil et al. (1981) study has been that the 25% of the cohort lost to follow-up were counted as survivors to the end of the study in 1978, thereby leading to a potential underestimation of the SMR. To avoid this potential bias, we adjusted the SMRs with the assumption that the rate of lung cancer in those lost to follow-up is the same as in those successfully followed up to the end of 1978. It is generally more difficult to verify survivorship in cohort studies than to identify deaths, thus it is likely that the actual cancer mortality would be less than that estimated by this method. The SMRs were therefore adjusted by multiplying by 1/0.75 because the initial slope was based on cases observed with 75% follow-up of the cohort.

The relative risk model (described earlier) is then utilized in the dose-response assessment. The SMR values from Table 3-2 were plotted against cumulative exposure (Fig. 3-1). A weighted least squares linear regression analysis using the expected numbers of lung cancer deaths as weights for each exposure category, and forcing an intercept of 1, produced a slope of 16.39 with an upper 95% CI of 20.02 (Table 3-3). This means that for the cohort under study, the upper confidence limit of the excess relative risk is 20.02 for every 1000 (mg/m³) years. The upper confidence limit is used to yield an estimate of risk that is unlikely to be exceeded.

When using SMRs, one can determine directly lifetime excess cancer risk for a population of workers exposed to nickel through inhalation, using the following model:

\[ R_x = R_0(1 + \beta) \]

where \( R_x \) represents the predicted risk to persons with exposure level \( x \), \( R_0 \) represents the background lifetime risk of dying from lung cancer without nickel exposure, and \( \beta \) represents the upper 95% CI of the slope from the linear relative risk model.

This model is also used to calculate the unit risk estimate (URE) for community exposures. However, a number of additional steps are neces-
Table 3.3. Dose-response calculations for lung cancer mortality based on the Ontario nickel refinery cohort study

<table>
<thead>
<tr>
<th>Parameter Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slope of lung cancer SMR versus exposure</td>
<td>16.39</td>
</tr>
<tr>
<td>Upper 95% confidence limit on slope</td>
<td>20.02</td>
</tr>
<tr>
<td>Unit cumulative exposure level in Figure 3-1</td>
<td>1000 (mg/m³ × years)</td>
</tr>
<tr>
<td>Exposure adjustment for 24 hr/day (8/24)</td>
<td>0.33</td>
</tr>
<tr>
<td>Adjustment for days per week (5/7)</td>
<td>0.71</td>
</tr>
<tr>
<td>Adjustment for weeks per year (48/52)</td>
<td>0.92</td>
</tr>
<tr>
<td>Adjusted units of cumulative exposure</td>
<td>216 (mg/m³ × years)</td>
</tr>
<tr>
<td>Equivalent lifetime (70-year) exposure level</td>
<td>5.08 mg/m³</td>
</tr>
<tr>
<td>Adjusted upper 95% CI on slope (20.02/3.08 mg/m³)</td>
<td>6.5</td>
</tr>
<tr>
<td>Background lifetime lung cancer mortality risk</td>
<td>0.049</td>
</tr>
<tr>
<td>Lifetime added risk for exposure to 1 μg/m³ (0.049 × 6.5/1000)</td>
<td>0.00052</td>
</tr>
</tbody>
</table>

Note: The calculations are explained in the text.

necessary in the linear extrapolation from the unit cumulative exposure of 1000 (mg/m³) years (Fig. 3-1) for the nickel cohort to a lifetime environmental exposure to 1 μg/m³. As shown in Table 3-3, adjustments are made for a 24-hour environmental exposure day versus an 8-hour workday, a 7 days per week environmental exposure versus a 5-day work week, and 52 weeks' environmental exposure versus a 48-week work week. The adjusted unit of cumulative exposure was determined to be 216 (mg/m³) years. Thus, for an environmentally exposed population, the excess relative risk would be 20.02 for each 216 (mg/m³) yr. In regulatory settings, environmental risk assessment is often based on a 70-year (lifetime) exposure duration. Therefore, the next adjustment in our extrapolation (i.e., 216 [mg/m³] yr/70 yr = 3.08 mg/m³) determines that an annual average lifetime exposure to 3.08 mg/m³ nickel yields an excess relative risk of 20.02. Next we will calculate the URE, or the excess cancer risk associated with a lifetime exposure to 1 μg/m³ nickel. The background lifetime lung cancer mortality risk for Canada in the years 1963 to 1978 was estimated to be 0.049, or 49 per thousand for males (WHO, 1966-1982). We derived this rate by dividing the total number of lung cancer deaths for the years 1963 to 1978 by the total number of deaths from all causes in those years. Using the model \( R_U = R_s(x + \beta) \), \( R_s = \) the background lifetime lung cancer mortality rate of 4.9% (or 0.049), \( x = 1 \mu g/m³ \), and \( \beta = \) the adjusted slope of 20.02/3.08 mg/m³, or 6.5 per mg/m³. Therefore, the excess number of lung cancers expected for a lifetime exposure to 1 μg/m³ is:

\[
\text{unit risk estimate} = 0.049 \times 1 \mu g/m³ \times 6.5/mg/m³ \times 1 mg/1000 \mu g = 3.2 \times 10^4
\]

or 320 additional lung cancer cases per million people exposed for a lifetime to 1 μg/m³ nickel in the air. Calculations of lifetime risks can be made more precise by adjusting for competing causes of death, as described by Gail (1975), but such adjustments usually have negligible effects.

The URE can then be used in the risk characterization phase of environmental risk assessment. For example, the lifetime risk of lung cancer for a community exposed to 0.01 μg/m³ nickel in air (a typical air concentration in urban settings), is calculated as follows:

\[
\text{excess risk} = 3.2 \times 10^4 (\mu g/m³)^3 \times 0.01 \mu g/m³
\]

or 3.2 extra cases of cancer for every million persons exposed to 0.01 μg/m³ for a lifetime.

The URE we have derived above for nickel exposure in Canada will be applicable to any other population if the lung cancer risks due to nickel are independent of background causes of lung cancer, particularly smoking. However, if the effects of nickel are synergistic with smoking, then the risks attributable to nickel would be partly dependent on the background rates of lung cancer in other populations of interest, because the effect of nickel exposure would be enhanced by smoking. Thus, if there was little smoking and therefore low background rates of lung cancer, then the risk of exposure to nickel would be lower in that population than if there were a lot of smoking and a high background rate of lung cancer. However, our assessment of the available evidence (Magnus et al., 1982) suggests that the effects of nickel and smoking are additive or close to it, and are therefore independent of background causes of lung cancer. Thus, the URE calculated here can be directly applied when characterizing the lung cancer risk to other populations.

**Sources of Uncertainty**

Quantitative cancer risk estimation for low-level environmental exposure to carcinogens involves many uncertainties. Nevertheless, systematic, logical, and informed approaches to decision making about carcinogens in the environment calls for quantitative assessments. The real problem with quantitative risk assessment is that findings are sometimes put forward with a degree of implied certainty that has no scientific basis. The estimate of cancer potency for environmental exposure levels must include zero in the range because it is possible that thresholds are present. Thus, the estimate that lifetime exposure to 1 μg/m³ nickel might increase cancer risks by as much as 320 per million should be qualified by stating that the
increased lung cancer risks may be zero, or they could fall somewhere in the range from zero to 320 per million, but are unlikely to be higher.

The major single source of uncertainty in the above risk assessment is the shape of the dose-response curve, which is based on sparse data. A linear relationship between exposure levels and relative risk was assumed because this fits the data well, and because we know of no evidence to suggest that the relationship is nonlinear.

The second major source of uncertainty involves the adjustment of workplace exposures to equivalent lifetime exposures. The assumption here is that relative risk at a given age relates to cumulative lifetime exposure, whatever the pattern of the exposure in terms of the age at which it was experienced. Thus it is implicitly assumed that someone experiencing continuous exposure from age 0 to age 50 would have the same relative risk as a worker with the same cumulative exposure measure also attained by age 50, but perhaps occurring over only 10 years and confined to the workday. On the one hand, it might be proposed that the lifetime exposure would carry a higher risk because it includes exposure as a child and it is possible that children are more susceptible to carcinogens in the environment. However, this seems unlikely on the basis of epidemiologic data showing very low lung cancer risks under the age of 40. In addition, reductions in cancer risks following cessation of exposure to carcinogens, for example the rapid reduction in lung cancer relative risk after stopping smoking (Office of Smoking and Health, 1982), provides evidence that equating cumulative lifetime risks to shorter-term workplace cumulative exposures should overestimate environmental carcinogen risks. Taken overall, it seems unlikely that the approach adopted leads to an underestimation of cancer risks from lifetime environmental exposures.

The third source of uncertainty involves the workplace exposure estimates themselves. Although it is possible exposures were higher or lower, it seems unlikely that actual exposures would be more than five times higher or five times lower than estimated. In the context of cancer risk assessment, the degree of uncertainty involved here is quite low. It might be noted that animal cancer bioassay studies have more accurate exposure data than occupational epidemiology studies, but the uncertainties in extrapolating risks to humans may involve an order of magnitude or more (Smith, 1988). Use of human data with a two- to fivefold uncertainty in exposure estimation involves less uncertainty overall.

**Meta-Analysis in Risk Assessment**

Meta-analysis is the structured and systematic qualitative and quantitative integration of the results of several independent studies. As stated earlier, meta-analysis may be a useful tool in the hazard identification and dose-response phases of health risk assessments based on epidemiologic data. Meta-analysis for which studies are the units of analysis, is different from pooled-data analysis, a method by which raw data from multiple studies of a single topic are combined in a single analysis. Although pooled-data analysis is preferable to meta-analysis in general (Freidrich, 1993), it is limited by the availability of raw data and is harder to do than meta-analysis. The focus of this section is the use of meta-analysis in hazard identification. Berin et al. (1993) provide a discussion of the meta-analysis of dose-response data.

The use of meta-analysis in assessing weak causal associations is controversial, especially for observational studies versus clinical trials. A recent group of articles has been published on this controversy (Shapiro, 1994 a, b; Petitti, 1994; Greenland 1994). Shapiro, a strong opponent of meta-analysis of observational studies, stated the opinion that “meta-analysis offers the Holy Grail of attaining statistically stable estimates for effects of low magnitude. In so doing, it ignores what is an absolute limit to epidemiological inference.” He proposed that meta-analysis of published non-experimental data should be abandoned.

Greenland (1994) argues that Shapiro’s criticisms apply to a prevalent and unsound form of meta-analysis that focuses on synthesis of study results into a single “conclusive” summary estimate. He believes the solution is to adopt a comparative approach in which meta-analysis is used as an aid in comparing studies and identifying patterns among study results. Greenland (1994) states that, “In the absence of ideal studies, there is a potential advantage of meta-analysis (and pooled-data analysis) over any single large study: When used to compare results from different studies meta-analysis can test hypotheses of biases. In contrast, inference from a single study is trapped within the framework of the study design.” He does agree, however, that there are problems with synthetic meta-analyses that ignore heterogeneity and report only a single fixed-effects summary.

Despite the controversy surrounding meta-analysis of observational studies, careful and critical application of appropriate meta-analytical techniques facilitates the quantitative exploration of heterogeneities and (where appropriate) syntheses of study results. Methods of quantitative meta-analysis have been reviewed in detail by Greenland (1987), Berlin et al. (1993), and DerSimonian and Laird (1986). The methodology of Greenland (1987) is emphasized here.

The first step in a well-designed meta-analysis is to define the objective—that is, formulate specific goals for the meta-analysis. In epidemiology, this usually means defining the measure of effects to be estimated, or defining the parameters to be estimated within some model for effects (Greenland, 1987). This step also involves the identification of study var-
variables to be included in the meta-analysis, including outcome and exposure.

The second step is to conduct a thorough literature search to gather all relevant published studies. It is possible that published studies are systematically different (more positive) from unpublished studies and therefore meta-analysis based on literature searches alone may lead to "publication bias." This source of bias is discussed further in the example presented later in this chapter. After one has gathered all the relevant published literature, and precisely specified the study exposure and outcome, the next step in the meta-analysis is to identify potential confounders. Some confounders may be generally recognized as so important that any report that fails to fully adjust for them will be immediately suspect (Greenland, 1987). An example of this is cigarette smoking in the study of asbestos and lung cancer.

At this stage of the meta-analysis, inclusion/exclusion criteria should be developed. Inclusion criteria should depend on the specific objectives of the analysis. Some of the variables on which inclusion criteria can be based are the study design, sample size, the outcome of interest, and whether or not the study is published (L'Abbe et al., 1987). Using these criteria, the reviewer should then list the papers to be included and excluded from the meta-analysis.

The next step in the meta-analysis is to record study characteristics for those studies to be included in the analysis. For example, the type of study (cohort, case-control), cohort size, record source, industry, number of deaths from the cause of interest (i.e., lung cancer deaths), the effect measure (i.e., SMR), and the confidence intervals. Extraction of study-specific effect measures and their standard errors from published studies may involve no more than copyng it out of the report. Greenland (1987) has reviewed computations that may be necessary to calculate standard errors should only confidence limits or $P$ values be presented rather than standard errors. One may also find that published results are partially or completely unadjusted for known or suspected important confounders, selection bias, and misclassification. Methods for making external adjustments to the effect measures are also described by Greenland (1987).

Once the review and reanalysis of the individual studies are completed, the statistical meta-analysis can be performed. The fundamental meta-analytic approach described by Greenland (1987) is a "fixed-effect" model and is based on weighted regression that treats each study result as the dependent variable with an accompanying weight. The weighted mean of the study results is then calculated. The appropriateness of the weighted mean as a meta-analytic summary of the effect under study depends on a very stringent homogeneity assumption. This assumption states that the studies are estimating the same value for the effect (i.e., "fixed effect"). In other words, after consideration of the extent of real effect and bias in each study, the studies should on average yield the same value, so that differences between the estimates are entirely due to random error. This will clearly not be true if the studies have different average doses, and if a greater dose causes a greater effect. In fact, one should regard any homogeneity assumption as extremely unlikely to be satisfied because of the differences in covariates, bias, and exposure variables among the studies (Greenland, 1987).

An important aspect of meta-analysis is to qualitatively and quantitatively analyze any heterogeneity that may be present. For multiple studies, the chi-squared test can be employed as basic statistical test of the homogeneity assumption; it is included in the example presented below. If a large amount of unexplained heterogeneity remains, one may consider turning to "random-effect" models (Colditz et al., 1995; Greenland, 1987; DerSimonian and Laird, 1986). With this approach, both random variation within studies and heterogeneity between studies is taken into account. When heterogeneity is present, the estimated confidence interval is more conservative (i.e., wider) than it would have been on the fixed-effect assumption. Criticism of the random-effect approach is that it is based on the assumptions that the studies are representative of some hypothetical population of studies and that heterogeneity between the studies can be represented by a single variance (Thompson and Pocock, 1991). Thompson and Pocock also note that undue weight may be given to small studies that themselves are the most likely to be influenced by publication bias. A number of investigators believe that a more useful approach is to focus on possible reasons for heterogeneity (Thompson and Pocock, 1991; Greenland, 1987). Colditz et al. (1995) have recently reviewed various approaches taken to identify, deal with, and interpret heterogeneity in meta-analysis of epidemiologic data and suggest methods that may be used in future studies. Berlin (1995) has supplemented the Colditz et al. (1995) review by commenting on the benefits of heterogeneity in meta-analysis of epidemiologic data. These and other issues arising from the application of meta-analytic methods will be discussed further in the example presented below.

The steps described earlier were undertaken in the following example of a simple meta-analysis using the fixed-effect model. We use as an example the possible association of silicosis and lung cancer. Although this example is not environmental, it nonetheless illustrates the basic principles of meta-analysis. This example is based on the more detailed meta-analysis performed by Smith et al. (1995).
Example: Is silicosis a risk factor for lung cancer?

Interpretation of the literature concerning the carcinogenicity of silica and an association between silicosis and lung cancer has been controversial. Studies among silicotics tend to demonstrate an excess risk of lung cancer but have been criticized because of possible selection and confounding biases. Other complications affecting the interpretation of the results arise from competing risks from other causes of death, such as silicosis and tuberculosis. In this example of meta-analysis in hazard identification, published data of lung cancer mortality among silicotics were combined to evaluate quantitatively the possible association of silicosis and an increased risk of lung cancer.

METHODS

This meta-analysis was conducted following general principles such as those outlined by L'Abbe et al. (1987). The epidemiologic literature was searched for all studies giving data concerning silicosis and lung cancer. When a study had been published in different articles, only the most recent report was included in the analysis.

The effect measures (relative risks, odds ratios, etc.) for lung cancer mortality among silicotics and their confidence intervals were extracted from each study for use in the meta-analysis. Missing confidence intervals were estimated by Byar's approximation (IARC, 1987) for cohort studies or by the "median-based method" described by Miettinen (1976) for case-control studies. If information on lung cancer mortality was not provided then total respiratory cancer mortality was used. When an SMR analysis used both national and regional populations to calculate expected deaths, the results based on regional disease rates were used in the meta-analysis.

The method of weighting by precision described by Greenland (1987) was used for the meta-analysis. The weight or inverse variance \( w = 1/SE^2 \) of each study was calculated using a standard error \( SE \) equal to the natural log of the ratio of the upper to lower 95% confidence intervals divided by 3.92 \( SE = \ln(\text{RR}_{\text{upper}}/\text{RR}_{\text{lower}})/3.92 \). The weight \( w \) was then multiplied by the natural log of the effect measure \( b \). A weighted mean or pooled summary \( \hat{b} \) was then calculated by dividing the sum of the weighted results \( \sum w b \) by the sum of the weights \( \sum w \).

Pooled summaries were estimated for all studies combined, and separately for cohort studies, case-control studies, and studies giving mortality odds ratios (MOR) or standardized incidence ratios (SIR). All pooled summaries were converted into rate ratios \( RR = \exp \hat{b} \) and 95% confidence intervals were estimated. A statistical test of heterogeneity was applied to all pooled summaries where \( \chi^2 = \sum w (b - \hat{b})^2 \) has a chi-squared distribution with degrees of freedom one less than the number of studies.

RESULTS

Twenty-nine studies were found in the published literature with data concerning lung cancer and silicosis. Information on the industry and record sources for the study participants, number of lung cancer deaths, effect measures and confidence intervals are detailed in Table 3-4. The study weights and any corrections are also given. Twenty-three of the 29 studies found were used in the meta-analysis.

Fourteen of the studies were cohort studies involving SMRs for lung cancer that ranged from 1.1 to 4.4. Armstrong et al. (1979) reported results for all respiratory cancers combined.

Four case-control studies included in the meta-analysis involved odds ratios ranging from 1.8 to 3.9 for lung cancer death. Additional studies included three cohort studies of lung cancer incidence (SIRs range from 1.7 to 2.9) and two studies that calculated mortality odds ratios (MOR=1.5 and 2.2).

Six studies were excluded from the meta-analysis because they suffered from biases that clearly underestimated or overestimated the risk of lung cancer for silicotics. For example, the proportionate mortality study of Rubino et al. (1990) underestimated the lung cancer mortality because the PMR was calculated without regard to competing causes of death for silicosis. The findings of the autopsy-based case-control studies of Hessel and Sluis-Cremer (1986), Hessel et al. (1990), and Hinrzu and Sluis-Cremer (1991) were also found to have underestimated the risk of lung cancer from silicosis. The controls in these autopsy studies may have had other diseases associated with silicosis.

Also excluded from the meta-analysis are two studies based on hospital records that gave SMRs of 5.0 (Chihotani et al., 1996) and 6.0 (Merlo et al., 1990). The findings from these studies are likely to be biased upward because silicotics with lung cancer are more likely to be admitted to hospitals than silicotics without lung cancer.

The pooled relative risk (RR) for all studies combined (excluding those with competing risk problems and studies based on hospital records) was 2.2 \( (P < .001) \) with a 95% confidence interval of 2.1–2.4 (Table 3-5). The highest pooled RR of 2.7 was found for the set of studies of lung cancer incidence (CI 2.3–3.2; \( P < .001 \)). Combination of cohort, case-control studies, and MORs yielded pooled RRs of 2.0 (CI 1.8–2.3; \( P < .001 \)), 2.5 (CI 1.8–3.2; \( P < .001 \)), and 2.0 (CI 1.7–2.4; \( P < .001 \)), respectively.
Table 3-4. Summary of studies of silicosis and lung cancer with effect measures, 95% confidence intervals (CI) and weightings (W) for meta-analysis

<table>
<thead>
<tr>
<th>Reference</th>
<th>Record Source/ Industry</th>
<th>Lung Cancer Deaths</th>
<th>Effect Measure</th>
<th>CI</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cohort Studies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amandus and Costello (1991)</td>
<td>Medical exam/mining</td>
<td>14</td>
<td>1.7</td>
<td>0.9-2.9</td>
<td>11.2</td>
</tr>
<tr>
<td>Amandus et al. (1991)</td>
<td>Medical exam/misc.</td>
<td>33</td>
<td>3.0</td>
<td>2.0-4.2</td>
<td>27.9</td>
</tr>
<tr>
<td>Armstrong et al. (1979)</td>
<td>Medical exam/mining</td>
<td>21</td>
<td>1.1</td>
<td>0.6-2.0</td>
<td>13.6</td>
</tr>
<tr>
<td>Carta et al. (1991)</td>
<td>Medical exam/misc.</td>
<td>22</td>
<td>1.5</td>
<td>0.8-2.0</td>
<td>13.3</td>
</tr>
<tr>
<td>Chen et al. (1992)</td>
<td>Silicosis registry/misc.</td>
<td></td>
<td>1.2</td>
<td>0.9-1.6</td>
<td>46.4</td>
</tr>
<tr>
<td>Finkelstein et al. (1987)</td>
<td>Compensation/mining  and surface workers</td>
<td>78</td>
<td>5.4</td>
<td>1.8-3.2</td>
<td>40.4</td>
</tr>
<tr>
<td>Infante-Rivard et al. (1989)</td>
<td>Compensation/misc.</td>
<td>93</td>
<td>5.5</td>
<td>2.8-4.5*</td>
<td>58.5</td>
</tr>
<tr>
<td>Mehnert et al. (1990)</td>
<td>Compensation/county</td>
<td>9</td>
<td>1.8</td>
<td>0.8-5.5</td>
<td>7.1</td>
</tr>
<tr>
<td>Neuberger et al. (1986)</td>
<td>Compensation/misc.</td>
<td>42</td>
<td>1.4</td>
<td>1.0-1.9</td>
<td>57.3</td>
</tr>
<tr>
<td>Ng et al. (1990)</td>
<td>Compensation/misc.</td>
<td>28</td>
<td>2.0</td>
<td>1.4-2.6</td>
<td>29.0</td>
</tr>
<tr>
<td>Puntoni et al. (1988)</td>
<td>Compensation/refractory brick</td>
<td>6</td>
<td>1.7</td>
<td>0.7-3.6</td>
<td>5.7</td>
</tr>
<tr>
<td>Tornling et al. (1990)</td>
<td>Pneumo. registry/misc.</td>
<td>9</td>
<td>1.9</td>
<td>0.8-3.6</td>
<td>6.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reference</th>
<th>Record Source/ Industry</th>
<th>Lung Cancer Deaths</th>
<th>Effect Measure</th>
<th>CI</th>
<th>W</th>
</tr>
</thead>
<tbody>
<tr>
<td>Westerholm et al. (1986)</td>
<td>Pneumo. registry/misc.</td>
<td>17</td>
<td>4.4</td>
<td>2.0-8.3</td>
<td>7.5</td>
</tr>
<tr>
<td>Zambon et al. (1986)</td>
<td>Compensation/misc.</td>
<td>49</td>
<td>2.3</td>
<td>1.7-5.0</td>
<td>47.6</td>
</tr>
<tr>
<td><strong>Case-control Studies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cocco et al. (1990)</td>
<td>Medical exam/misc.</td>
<td>15</td>
<td>2.4</td>
<td>1.0-6.2</td>
<td>4.6</td>
</tr>
<tr>
<td>Lagorio et al. (1990)</td>
<td>Compensation/misc.</td>
<td>15</td>
<td>3.9</td>
<td>1.8-8.3</td>
<td>6.6</td>
</tr>
<tr>
<td>Mastrangelo et al. (1988)</td>
<td>Compensation/misc.</td>
<td>50</td>
<td>1.8</td>
<td>1.1-2.8</td>
<td>17.6</td>
</tr>
<tr>
<td>Steenland &amp; Beaumont (1986)</td>
<td>Silicosis on death certificate/granite</td>
<td>26</td>
<td>3.2</td>
<td>1.6-6.4*</td>
<td>8.3</td>
</tr>
<tr>
<td><strong>Standardized Incidence Ratios</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chua et al. (1991)</td>
<td>Silicosis registry/misc.</td>
<td>9</td>
<td>2.9</td>
<td>0.9-5.9</td>
<td>7.1</td>
</tr>
<tr>
<td>Partner et al. (1994)</td>
<td>Silicous registry/misc.</td>
<td>101</td>
<td>2.9</td>
<td>2.4-5.5</td>
<td>107.5</td>
</tr>
<tr>
<td>Sherson et al. (1991)</td>
<td>Medical exam/ foundry</td>
<td>11</td>
<td>1.7</td>
<td>0.9-3.1</td>
<td>10.0</td>
</tr>
<tr>
<td><strong>Mortality Odds Ratios</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foransiere et al. (1989)</td>
<td>Compensation/ ceramic</td>
<td>64</td>
<td>1.5</td>
<td>1.1-1.9</td>
<td>51.4</td>
</tr>
<tr>
<td>Schuler and Ruttner (1986)</td>
<td>National Accident Ins. Fund/misc.</td>
<td>186</td>
<td>2.2</td>
<td>1.8-2.7*</td>
<td>93.5</td>
</tr>
</tbody>
</table>

* Confidence: intervals calculated based on Byar's approximation.

The results of the test of homogeneity were statistically significant (i.e., the studies are considered heterogeneous) for all studies combined and for the set of cohort studies (Table 3-5). A heterogeneity $P$ value of 0.3 was estimated for the four case-control studies. The three studies of lung cancer incidence yielded a $P$ value of approximately 0.1.

**DISCUSSION**

The results of this meta-analysis suggest that there is approximately a two-fold increase in the risk of lung cancer for patients with silicosis. The pooled RRs are consistent across study type. When considered individually, all of the studies found in the literature demonstrated effect measures greater than 1.0. Fifteen of these studies reported confidence intervals that exclude 1.0. Consequently, it is extremely unlikely that the increased lung cancer risk is attributable to chance.

One potential bias affecting the relationship between silicosis and lung cancer is confounding by cigarette smoking. Data on smoking have not been consistently available, and some studies relied on indirect methods of controlling for smoking (Ng, 1990; Forastiere et al., 1989; Merlo et al., 1990). However, four studies included in the meta-analysis did adjust for smoking in their analysis of lung cancer risk among silicotics (Amandus and Costello, 1991; Amandus et al., 1991; Coccoli et al., 1990; Lagorio et al., 1990). The smoking-adjusted effect measures were found to be higher than the unadjusted results for all four studies. These findings suggest that the increased risk of lung cancer among silicotics is not attributable to smoking. In fact, there is little consistent evidence that smoking is a risk factor for silicosis, and therefore confounding by smoking would not be expected, or would at best be a very weak confounder in the relationship between lung cancer and silicosis.

Another question concerns whether or not silicosis can lead to lung cancer in the absence of smoking. Elevated lung cancer risks were apparent in two studies that calculated expected deaths from lung cancer based on nonsmokers (Amandus et al., 1991; Mastrapasqua et al., 1988). An SMR of 8.6 (CI 5.6–20.5) was calculated in the study by Amandus et al., and an OR of 5.3 (CI 0.5–43.5) was obtained in the Mastrapasqua et al. study. Thus, the overall evidence supports increased risks of lung cancer in silicotics who do not smoke.

Other possible biases might result from the choice of external referent populations based on other lifestyle or socioeconomic differences between occupational groups and national populations. In addition, studies based on compensation for silicosis could be biased toward a silicosis–lung cancer relation if diagnosis of lung cancer affected detection of silicosis. Selection bias might also result if cases of silicosis were determined by

<table>
<thead>
<tr>
<th>Cohort</th>
<th>All Studies</th>
<th>Case-control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted sum (Z)</td>
<td>54.3</td>
<td>206.5</td>
</tr>
<tr>
<td>Weighted sum (Z)</td>
<td>692.3</td>
<td>385.4</td>
</tr>
<tr>
<td>Summary $\hat{R}$</td>
<td>6.9</td>
<td>6.8</td>
</tr>
<tr>
<td>Summary $\hat{R}$</td>
<td>6.9</td>
<td>6.8</td>
</tr>
<tr>
<td>Confidence interval (CI)</td>
<td>4.1–11.3</td>
<td>4.1–11.3</td>
</tr>
<tr>
<td>Z statistic (Z)</td>
<td>11.7</td>
<td>11.7</td>
</tr>
<tr>
<td>Homogeneity chi-squared</td>
<td>20.9</td>
<td>20.9</td>
</tr>
<tr>
<td>Degrees of freedom</td>
<td>22</td>
<td>22</td>
</tr>
</tbody>
</table>

$\hat{R} = \log (\text{risk measure})$
voluntary medical examination. Smith et al. (1995) discuss these potential biases in greater detail.

The studies of lung cancer risk among silicotics have also been criticized because they often include workers from industries where exposure to other lung carcinogens would have occurred (i.e., PAHs in foundries; radon and arsenic in mines). However, elevated lung cancer risks have been demonstrated in silicotics from granite and stone industries where the principal risk factor is exposure to silica and exposure to other potential carcinogens is unlikely (Finkelstein et al., 1987; Infante-Rivard et al., 1989; Mehner et al., 1990; Zambon et al., 1986; Steenland and Beaumont, 1986; Kurppa et al., 1986).

Publication bias is a potential limitation in any meta-analysis. Although a tendency to report and publish positive results has been well documented, results of large studies generally approximate the results of all studies published and unpublished (Felson, 1992). Moreover, large studies are generally published whether or not their results are "positive" (Begg and Berlin, 1989). Of the 29 studies found in the peer-reviewed literature, seven studies examined cohorts of over 1000 silicotics (Chen et al., 1992; Finkelstein et al., 1987; Infante-Rivard et al., 1989; Ng et al., 1990; Zambon et al., 1986; Schuler & Rutten, 1986; Chiyotani et al., 1990). All of the lung cancer relative risk estimates were greater than 1.0. Six of the seven studies were included in the meta-analysis. The pooled effect measure for these studies is 2.2 (95% CI, 2.0–2.5). Ten of the 29 studies reported results for 50 or more lung cancer deaths or cases (Finkelstein et al., 1987; Infante-Rivard et al., 1989; Mastrangelo et al., 1988; Partanen et al., 1994; Forastiere et al., 1989; Schuler & Rutten, 1986; Rubino et al., 1990; Hessel and Sluijs-Cremer, 1986; Hessel et al., 1990; Hnizdo and Sluijs-Cremer, 1991). Six of these studies were included in the meta-analysis. The pooled RR was estimated at 2.5 (95% CI, 2.2–2.7). Thus, publication bias is unlikely in this meta-analysis.

Heterogeneity among studies is another problem in many meta-analyses, as discussed earlier. Any group of epidemiologic studies is likely to be heterogeneous by virtue of differences in exposure level, in study design, variation in background rates and differences in length of follow-up, case identification, and so on. The advantage of the weighting method applied in the present analysis is that effect measures from studies with small sample sizes contribute less to the pooled estimate. Possible reasons for heterogeneity particular to studies among silicotics include differences between countries in defining silicosis and possible differences in disease detection methods, patterns of smoking, and choice of referent population. The form of crystalline silica (for example, quartz and cristobalite) to which individuals are exposed would also vary. Because the fibrogenic potential of the various forms differs, so might the potential carcinogenicity. The sources of heterogeneity in this meta-analysis are analyzed in greater detail by Smith et al. (1995).

Finally, the basic hypothesis of this review, that silicosis may lead to lung cancer, is biologically plausible. Silica-induced carcinogenesis has been demonstrated based on studies in the rat (Dagle et al., 1986; Groth et al., 1986; Holland et al., 1983; Holland et al., 1986; Muhle et al., 1989; Spiethoff et al., 1992). However, neither fibrosis nor carcinogenesis has been demonstrated in silica-exposed hamsters (Holland et al., 1983; Renne et al., 1985; Saffioti, 1992). Silica has been shown to induce fibroblastic lesions in the mouse although the carcinogenicity of silica has not been adequately tested in this species (Holland, 1990).

The results of this meta-analysis demonstrate that there are, indeed, increased risks of lung cancer among persons diagnosed with silicosis. It remains unclear whether silicosis itself is involved in lung cancer etiology or whether it is instead an indicator of heavy exposure to silica.

CONCLUSIONS

Environmental risk assessment is the quantification of potential adverse health effects of human exposure to environmental hazards. Good epidemiologic studies, when available, are generally considered to be the best source of data for use in risk assessment. One advantage is that they provide direct evidence for carcinogenic or other health effects in humans, thus avoiding the uncertainty of interspecies extrapolation. However, epidemiologic studies do not show cause-effect or dose-response relationships with the same ease as do experimental studies in animals. Clearly, more can be learned from available epidemiologic studies by a careful, critical, and comprehensive review. Because of the rapid growth of epidemiology over the past decades, the traditional narrative literature review is not always the simplest method of summarizing the results of a collection of studies. Meta-analysis, if appropriately conducted, is an alternative approach to the quantitative and qualitative analysis of a collection of epidemiologic study results, and it can be a valuable tool in the risk assessment process.

References


Shapiro S. 1994b. Is there is or is there ain’st no baby: Dr Shapiro replies to Drs Peititi and Greenland. Am J Epidemiol 140(9): 788–791.


