Introduction

Cigarette smoking is a major cause of cancer of the lung, larynx, oral cavity, and esophagus and is a contributory factor for cancer of the kidney, urinary bladder, and pancreas (US DHHS 1982). These cancers will cause 278,700 of the estimated 910,000 new cancer cases in the United States during 1985 (ACS 1985), or 30.6 percent of the cancers occurring in the United States other than skin cancer. Exposures to agents in the workplace other than cigarette smoke will also cause some of these new cancers, and a number of cancers will result from the combined effects of cigarette smoking and carcinogenic exposures in the workplace.

The role that cigarette smoking plays in causing these cancers is well established and extensively documented (US DHHS 1982). The role that occupational agents play in the development of these same cancers continues to emerge as the effects of more agents are examined both in the laboratory and in the workplace. However, cigarette smoking by exposed workers makes it difficult to separate the effects of smoking from the effects of occupational agents for cancers of sites causally linked to cigarette smoking. For some agents, such as asbestos, both the large numbers of people exposed and the magnitude of the increased cancer risk have allowed a careful examination of the relative contributions of cigarette smoking and the workplace exposure. For most agents, the data are more limited. Nevertheless, protection of workers requires that regulatory decisions be made about individual workplace exposures, even in the face of limited data. In assessing the effects of workplace exposures, consideration must be given to the interactions of smoking with agents that increase risk and to the bias introduced into studies of occupational groups by confounding effects of cigarette smoking. This chapter discusses the nature and measurement of interactions between smoking and occupational exposures and the sources and control of confounding of smoking and occupational exposures. It is not intended to be a comprehensive discussion of the epidemiologic methods used to evaluate workplace exposures, but rather a discussion of how smoking behavior in the workforce can effect the evaluation of occupational exposures. The data on smoking and specific occupational exposures are presented in later chapters of this Report. The discussion of these issues is intended to aid in the design and interpretation of studies of occupational exposure and not to criticize those studies in which smoking could not be completely addressed.

Lung Cancer Death Rates and Smoking

A detailed discussion of the causal relationship between cigarette smoking and the cancers is provided in an earlier Report in this

series (US DHHS 1982) and is not repeated here. However, the relationship between smoking and lung cancer is briefly described, as a framework for the discussion of interaction and confounding in subsequent sections of this chapter. Lung cancer was chosen as an example because of its strong link to smoking and because it is the greatest cause of cancer death in both men and women (ACS 1985).

Lung cancer will cause an estimated 125,600 deaths in 1985 (ACS 1985): 87,000 men and 38,600 women. For men, this represents more than 8 percent of all deaths. Current U.S. age-specific lung cancer death rates increase with age into the late seventies age range and then decline. However, when death rates for any given birth cohort of men are examined (Figure 1), there is no decline in death rates at the older ages. This difference between the cross-sectional mortality statistics and the cohort data is generally attributed to differences in the smoking habits of successive birth cohorts of men (and women) during this century. This Report's chapter on smoking patterns in the U.S. population also carefully documents that cigarette smoking is not uniformly distributed in the U.S. population, but rather varies considerably with both age and occupation. This nonuniform distribution of smoking patterns introduces much of the difficulty in controlling for smoking in occupational studies.

The relationships among age, lung cancer death rates, and number of cigarettes smoked per day, derived from the mortality study of U.S. veterans (Kahn 1966), are presented in Figure 2. The risk associated with smoking is a function of both the intensity of smoking, as measured by number of cigarettes smoked per day and depth of inhalation, and the duration of smoking as measured by age and age of initiation.

The lung cancer mortality ratios derived from the American Cancer Society (ACS) study of 1 million men and women (Hammond 1966) for smokers compared with nonsmokers, stratified by age and by number of cigarettes smoked per day, depth of inhalation, and age of initiation are presented in Table 1. In general, the mortality ratios are greater in the older age groups and increase with increasing dosage measure within each age strata. The data demonstrate that within the broader category of smokers a substantial variation in risk (up to fivefold) occurs between the different levels of dose and duration of smoking. The variation in mortality ratios for each isolated measure in Table 1 almost certainly overestimates the independent contribution of that measure to the actual risk, owing to correlation among the measures of number of cigarettes smoked per day, depth of inhalation, and age of initiation. For example, those who begin to smoke at a young age also smoke more cigarettes per day (Shopland and Brown 1985). However, it is unlikely that this correlation among dosage and duration measures explains all of the variation in mortality ratios with the isolated measures; therefore, it

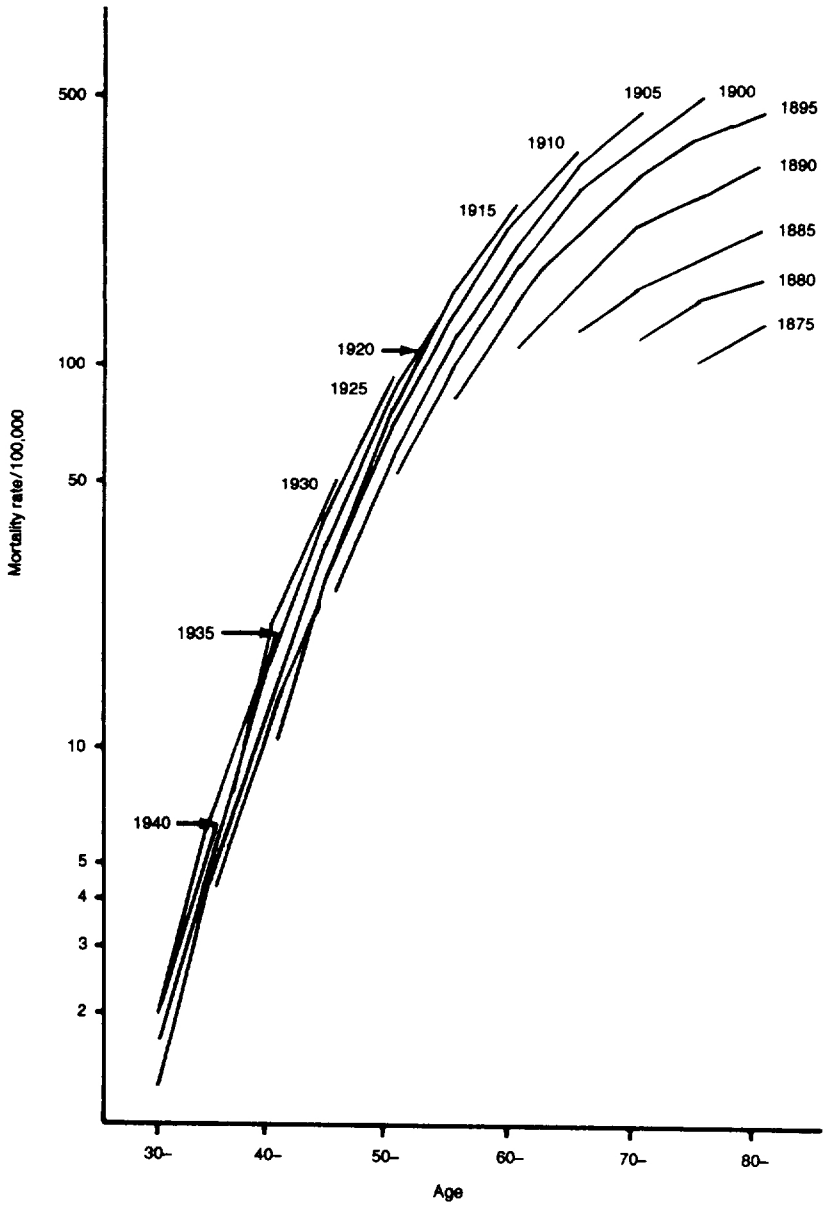

FIGURE 1.—Age-specific mortality rates for cancer of the bronchus and lung, by birth cohort and age at death, men, United States, 1950–1975

SOURCE: Data derived from McKay et al. (1982).

is reasonable to expect that the accuracy of lung cancer risk estimates for a population would improve with the inclusion of a

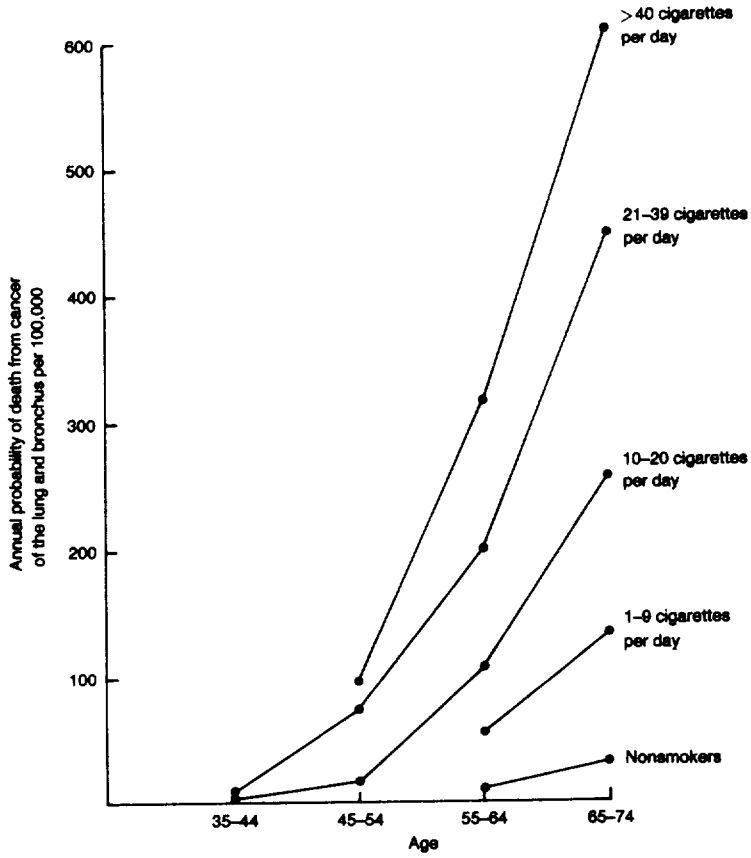

FIGURE 2.—Death rates from cancer of the lung and bronchus in nonsmokers and smokers of various numbers of cigarettes per day

SOURCE: Kahn (1966).

measure of smoking prevalence, a measure of smoking intensity, a measure of smoking duration, and a measure of the duration of cessation for former smokers.

Interactions Between Cigarette Smoking and Occupational Exposures

Interactions between cigarette smoking and occupational exposures may be examined in the context of a biological process, as a statistical phenomenon, or as a problem in public health and individual decisionmaking (Rothman et al. 1980; Saracci 1980; Siemiatycki and Thomas 1981). In each of these contexts the

TABLE 1.—Number of lung cancer deaths (men), age-standardized death rates, and mortality ratios, by current number of cigarettes smoked per day, degree of inhalation, and age began smoking, by age at start of study

Smoking characteristics	Age 35–54			Age 55–69			Age 70–84			All ages, 35–84		
	Number of deaths	Death rate	Mortality ratios	Number of deaths	Death rate	Mortality ratios	Number of deaths	Death rate	Mortality ratios	Number of deaths	Death rate	Mortality ratios
Current number of cigarettes a day												
1–9	9	38	6.17	12	68	3.53	5	134	5.32	26	56	4.60
10–19	15	24	3.90	57	168	8.77	10	243	9.62	82	90	7.48
20–39	138	58	9.37	216	264	13.82	27	446	17.62	381	159	13.14
≥40	26	47	7.67	50	334	17.47	6	754	29.84	82	201	16.61
Degree of inhalation												
None or slight	19	29	4.75	87	203	10.60	14	193	7.65	120	102	8.42
Moderate	114	52	8.48	177	224	11.72	20	401	15.88	311	138	11.45
Deep	55	55	9.00	73	266	13.93	13	638	25.26	141	173	14.31
Age began cigarette smoking												
≥25	5	17	2.77	12	65	3.39	3	85	3.38	20	39	3.21
20–24	31	36	5.83	72	212	11.11	7	306	12.11	110	118	9.72
15–19	112	54	8.71	176	250	13.06	27	490	19.37	315	155	12.81
<15	35	79	12.80	57	302	15.81	9	424	16.76	101	183	15.10
Never smoked regularly	11	6		27	19		11	25		49	12	

NOTE: Mortality ratios are based on death rates carried out to one more significant figure than shown.
SOURCE: Hammond (1966).

concepts are applied somewhat differently, and confusion results when a move from one context to another is attempted without consideration of these differences in application. Biological interaction refers to the presence of one agent influencing the form, availability, or effect of a second agent, and includes physical interaction such as the adsorption of carcinogens to particulates in inspired air, process interactions such as the induction by one agent of an enzyme system capable of converting a second agent into a carcinogenic metabolite, and outcome interactions such as the number of tumors produced by separate and combined exposures in an animal exposure system. Statistical interaction refers to a departure from the mathematical model used to assess the effects of the exposure variables. The model being tested may be additive, multiplicative, or some other form; the outcome of interest may be death rates, relative risks, or other outcome measures; the independent variables may be intensity of exposure, duration of exposure, a combination of intensity and duration (e.g., pack-years), or a logarithmic or other transformation of these measures. Public health interaction usually refers to the presence or level of one agent influencing the incidence, prevalence, or extent of disease produced by a second agent. An exposure to two agents that resulted in a multiplicative effect on lung cancer death rates might show no interaction using a multiplicative statistical model, but might show a profound interaction in terms of public health and a variety of interactions within the biologic system under consideration (i.e., human carcinogenesis).

Biologic Interactions

The transformation of normal lung tissue into a clinically manifest lung cancer is a complex, incompletely understood process that is generally assumed to require multiple inheritable changes within the cell (Armitage and Doll 1961; Day and Brown 1980). Although cellular changes are assumed to be requisite for carcinogenesis, phenomena taking place outside the cell may influence carcinogenesis. Cigarette smoke and occupational agents may potentially interact by influencing the fraction of inhaled carcinogen deposited and retained in the lung, the rate of metabolic activation of a procarcinogen into a carcinogenic metabolite, the transfer of agents across mucosal and cellular boundaries, the vulnerability of the cell to carcinogenic change (by increasing the rate of cell replication), or the transformation of the cellular DNA. In addition, cellular DNA repair, humoral or metabolic factors influencing tumor growth, and immunologic recognition or destruction of tumor cells are processes that may influence tumor manifestation and may be affected by occupational exposures and cigarette smoke. A detailed discussion of chemical carcinogenesis is beyond the scope of this chapter and is

provided elsewhere (Weinstein 1985; Farber 1982); however, this chapter explores some potential sites of biological interaction between occupational exposure and cigarette smoke to illustrate the biologic interactions that may take place.

Cigarette smoking and occupational exposures may interact through effects of smoking on the dose of the carcinogen that reaches the cell. Long-term exposure to cigarette smoke impairs mucociliary clearance (US DHHS 1982) and could alter the dose of an occupational agent retained. Carcinogens may adsorb to particulates in smoke or to environmental dusts (Natusch et al. 1974; Mossman et al. 1983), resulting in a higher fractional retention or different distribution in the lung. The adsorption to dust may also facilitate or inhibit transport of carcinogens through the mucus layer. Cigarette smoke has been shown to increase epithelial permeability in the tracheobronchial tree (Simani et al. 1974); the effect may increase the exposure of the underlying cell to an occupational agent.

Another potential site of biologic interaction is the metabolic activation of a carcinogen. A number of agents, including the polycyclic aromatic hydrocarbons in cigarette smoke, undergo chemical transformation within the body to metabolites that are considered to be active carcinogens (Gelboin and Tso 1978a, b). The majority of known conversions occur through the mixed function oxygenase system predominately located in the microsomal fraction of the cell. A number of constituents of cigarette smoke have been shown to induce this enzyme system (US DHEW 1979), and its activation may increase the rate of biologic activation of procarcinogens in the worksite. Cigarette smoking also alters the cellular composition of the lung, increasing the number of neutrophils and activated macrophages in the lung (US DHHS 1984); these cells may also play a role in the metabolic transformation of occupational agents.

Much of the consideration of interactions between smoking and occupational exposures has centered on interactions that might influence the response of the cell rather than the "dose" of carcinogen (Siemiatycki and Thomas 1981; Rothman et al. 1980; Rothman 1974, 1978; Walter and Holford 1978). In a widely accepted conceptual model, the process of malignant transformation of a cell into a cancer is considered to be a multistage process requiring multiple inheritable changes (Armitage and Doll 1961; Day and Brown 1980). Individual agents may initiate or promote the process of carcinogenesis. Initiation is thought to be at least a two-stage process that requires cell division before becoming irreversible (Farber 1982). Promotion describes the process by which an agent encourages an initiated tissue to develop focal proliferation. A tumor initiator may exert its effect through a brief exposure, whereas a tumor promoter usually requires repetitive contact with initiated

tissue to exert its effect. Cigarette smoke is known to contain a number of compounds that act as tumor initiators and promoters (US DHHS 1982); occupational exposures reflect a similar range of agents. Tumor promoters in smoke may influence the effects of exposure to tumor initiators in the workplace and thus increase the number of cancers that occur, and the presence of tumor initiators in smoke may allow the expression of a tumor promoter in the worksite.

The process of carcinogenesis is frequently modeled as a multistep process in which each succeeding step can occur only in those cells that have undergone the preceding step (Armitage and Doll 1961; Day and Brown 1980). In this model, agents may influence one (or more) of these steps, and therefore may have an effect early or late in the carcinogenic transition. Because the later steps in the process can occur only in cells that have undergone the changes of earlier steps, agents that act at separate steps may have multiplicative effects. For example, an agent that results in a fourfold increase in the rate of transition from a hypothetical step 1 to step 2 in the carcinogenic process would result in a fourfold increase in the number of malignant transformations by increasing the number of cells available for step 2 and subsequent steps. Similarly an agent that tripled the rate of transition from step 2 to step 3 would triple the number of malignant transformations. However, exposure to both agents would provide a fourfold (300 percent) increase in the number of cells available for transition from step 2 to step 3 as well as a threefold (200 percent) increase of the rate of transition from step 2 to step 3, with a resultant twelvefold (1,100 percent) increase in the number of malignant transformations. Therefore, the effect of the combined exposure on number of malignant transformations (1,100 percent) would be greater than the sum of the effects of independent exposures (300 percent plus 200 percent).

A similar phenomenon may occur with cigarette smoke and an agent that has an independent and additive effect as an initiator of carcinogenesis. The additive effects on tumor initiation may appear as a multiplicative effect on tumor occurrence because of the action of the tumor promoters in cigarette smoke. The tumor promoters in smoke may act on the cells initiated by an occupational agent, as well as on the cells initiated by smoke, to increase the number of the cells that become cancers. The number of tumors produced by a combined exposure could then be greater than the sum of the numbers of tumors produced by the individual exposures separately.

Two additional mechanisms by which cigarette smoking and occupational exposures may interact are by alterations in the immunologic surveillance for cancers and by increasing the frequency of cell division. Differences in the number, type, and function of cellular components of the immune system have been demonstrated

between smokers and nonsmokers (US DHHS 1984) and among workers exposed to occupational agents (see other chapters of this Report). The potential for these differences to influence the rates of clinically manifest cancers (either positively or negatively) is an issue of considerable interest. The increase in cell turnover in the respiratory tract in response to the acute toxic and inflammatory effects of cigarette smoke, or of occupational exposures, may also influence cancer rates, as it is believed that cells are more vulnerable to carcinogenic changes during periods of replication.

This discussion is intended to illustrate the kinds of biologic interactions that might occur between smoking and occupational agents and not to be a complete description of either the carcinogenic process or the sites of potential interaction.

Statistical Interaction

Statistical interaction refers to departure from a mathematical model in assessing the main effects of independent variables; its presence is often evaluated by the addition of an interaction term to the independent variables (Siemiatycki and Thomas 1981; Blot and Day 1979; Saracci 1980). With this approach, the presence of interaction is dependent on the model being used (Rothman 1974; Kupper and Hogan 1978). For example, a multiplicative effect can be adequately modeled without an interaction term on a log scale, but requires an interaction term on an additive scale. In this section, an additive model for the effects of two exposures assumes that the combined exposure produces an effect equal to the background rate plus the sum of the increases from the background rate of the two exposures experienced separately. In a multiplicative model, combined exposure results in an effect equal to the product of the effects produced by the separate exposures.

The following example illustrates this terminology and demonstrates the dependence of statistical interaction on the selected model. Assuming that two agents independently increase the risk of lung cancer and that the separate exposures result in a fivefold and tenfold increase in risk, respectively, if exposure to both agents produces an eightfold increase in risk, there is negative interaction (protective effect) in the additive and the multiplicative models. A combined risk of 14 indicates no interaction in an additive model, but a negative interaction in a multiplicative model; a risk of 30 is a positive interaction with an additive model and negative with a multiplicative model; a risk of 50 is a positive interaction with an additive model and no interaction with a multiplicative model; and a risk of 60 is a positive interaction with both models.

This example illustrates the critical dependence of tests for interaction on the mathematical model that is selected. Ideally, the choice of a model is based on biological considerations and not on

statistical convenience. For example, if the potential interaction of two initiators is being examined, an additive model should be used. The use of a multiplicative model may result in the demonstration of a negative interaction.

When applied to the multistage biologic model of carcinogenesis, independent actions at the same step would yield additive effects and actions at separate steps would yield multiplicative effects (Siemiatycki and Thomas 1981; Walter and Holford 1978). This progression from the biologic model to the statistical effect is easily defended; however, it is less clear that the reverse progression is valid, particularly in epidemiologic studies. The demonstration of an additive effect on lung cancer death rates does not necessarily imply that the two agents are acting at the same point in the carcinogenic process, nor does a multiplicative effect guarantee action at separate steps. As should be evident from the discussion of biologic interaction, cigarette smoke may interact with occupational agents at points external to the cell, and smoke consists of a variety of agents with different carcinogenic effects. The complex biologic processes that underlie the exposure–disease relationships evaluated in epidemiological studies limit the inference from the results of statistical modeling to biological mechanisms.

Rothman (1974) and Hogan and colleagues (1978) described methods of quantifying the magnitude of statistical interaction, and Kupper and Hogan (1978) described the detection of interaction in cohort and case–control studies. This Report's chapter on the evaluation of chronic lung disease also discusses the concepts of interaction and its measurement in studies of outcomes that are continuous (i.e., lung function measures) rather than binary (i.e., presence or absence of lung cancer).

In the simplest analytical problem, departure from additivity can be readily assessed when a population has two exposures, the rates in the presence of each individual exposure are known, and the rates in the presence and absence of both are known. If the relative risk (RR) in the absence of exposure is set equal to 1, then the ratio of the rate in the population with only one of the exposures to the rate in the population with neither exposure is the RR associated with the exposure. Correspondingly, the ratio of the rate in the population with both exposures over the rate in the population with neither exposure is the RR associated with combined exposure. The magnitude of the interaction can then be estimated by the ratio of the increase in rate with combined exposure (the RR of combined exposure minus 1) over the sum of the increases from the unexposed rate produced by the single exposures ($(RR_a-1)+(RR_b-1)$). The confidence interval around this estimate of interaction can also be estimated (Rothman 1974) as a measure of its statistical significance. More complicated estimates of the magnitude of interaction are

necessary when the rate in the unexposed population is unknown, when the rate of the disease being measured is high in the general population, and when case–control analyses are being performed (Rothman 1974; Hogan et al. 1978). In general, the size of the population needed to test for interaction between two exposures is considerably larger than the size of the population needed to establish statistically significant effects for the separate exposures.

Both case–control and cohort data can be analyzed with approaches that involve stratification (Kleinbaum et al. 1982; Rothman and Boice 1979). The data are separated into strata defined by levels of the occupational exposure and of cigarette smoking. By combining the information within the separate strata, summary measures can then be calculated that estimate the independent effects of the variables and describe their interaction. Although stratified analysis can be readily performed, its application is frequently limited by the number of available subjects, both in the entire study and within specific strata. For example, if an investigator designates four levels of exposure to an occupational agent and classifies smokers as currently smoking, previously smoking, or never smoking, twelve separate exposure categories are created. If age, sex, and race must also be considered, stratified analysis may be feasible only if the number of subjects is extremely large.

Statistical modeling represents an alternative that is less compromised by smaller sample sizes and that provides greater flexibility for controlling confounding and for testing for interaction. Modeling refers to the specification of a particular mathematical relationship between the outcome variable, e.g., the occurrence of lung cancer, and the variables representing the exposures of interest, e.g., cigarette smoking and an occupational agent. Statistical methods describe the adequacy of the model for the data and provide estimates of the effects of the exposure variables. Modeling can be performed with the programs available in most conventional statistical packages, but some special applications may require customized software.

In analyzing data on the effects of occupational exposures in populations with a high prevalence of smoking, modeling facilitates the control of confounding by smoking; multiple variables that characterize smoking, such as duration, daily amount, and depth of inhalation, can be entered simultaneously into the model. Further, if the cumulative exposures to the occupational agent and to cigarette smoke are temporally correlated, modeling may more satisfactorily separate their effects, in comparison with stratified analysis.

A recent report by Whittemore and McMillan (1983) illustrates the application of modeling to occupational data. These investigators analyzed data collected in the U.S. Public Health Service study of Colorado Plateau uranium miners, a prospective cohort study of

mortality in relationship to exposure to radon daughters in the mines. Their analysis assessed exposure to radon daughters and cigarette smoking as risk factors for lung cancer. To assess the joint effects of smoking and radiation, they developed and contrasted additive and multiplicative models. They found that the multiplicative model fit the data better than the additive. Of the alternative multiplicative models, giving the highest likelihood of the data was a linear function of the variables for smoking and radon daughter exposure. Whittemore and McMillan then used this multiplicative model to assess the effects of age and birth cohort. This analysis complemented the conventional cohort methods that had been applied previously to the data (Lundin et al. 1971; Archer et al. 1976).

Most conventional forms of modeling assume either an additive or a multiplicative relationship between the independent effects of the variables representing the exposures. Case–control data are most often analyzed with the multiple logistic model (Breslow and Day 1980; Schlesselman and Stolley 1982), although alternatives have been described (Walker and Rothman 1982; Breslow and Storer 1985). The multiple logistic model is multiplicative; the risk of disease from multiple exposures is obtained as the product of the risks from the individual exposures, in the absence of interaction among the exposures. A variety of approaches have been described for the modeling of data from cohort studies (Breslow et al. 1983; Breslow 1985). These models may be developed as additive or as multiplicative or on other scales.

In developing a model, confounding is controlled by introducing variables for the potentially confounding exposures. Statistical interaction among the variables is tested by entering terms formed as their product or by running the model within groups of subjects separated by their classification on one of the exposure variables. When a product term is entered into a model to test for interaction, the presence and extent of interaction is indicated by the coefficient calculated for the product term. Most modeling techniques also supply a test of statistical significance for the coefficient, under the null hypothesis that its value is zero. Such a test of statistical significance may not be very powerful (Greenland 1983), and the coefficient may suggest an interaction of potentially important magnitude, although it does not reach statistical significance at conventional levels.

The presence of statistical interaction between two variables demonstrates that their effects are interdependent, as assessed by the specific statistical model (Rothman et al. 1980). Statistical interaction does not necessarily imply biological interaction. In fact, the interpretation of interaction hinges on the scale on which it is measured; the choice of the statistical model may determine whether

interaction is present or absent, synergestic or antagonistic (Greenland 1979; Rothman et al. 1980). If possible, the choice of model should be based on biological considerations. For malignancy, the results of modeling may be interpretable within the conceptual framework supplied by the theory that carcinogenesis is a multistep process (Armitage and Doll 1961; Day and Brown 1980).

Public Health Interactions

From a public health perspective, an interaction occurs when the number of individuals injured, or the extent of the injury, with combined exposure exceeds that expected from the sum of the background rate and the differences between the background rate and the rates with the individual exposures. Public health interactions can be considered a case of statistical interaction in which both the model being tested and the outcome measurement scale being used are defined by their ability to assess the contribution of a given agent to the disease burden in society. When a positive interaction occurs in this definition, the term "synergism" should be used. The model used to examine interactions is often further specified by the importance of considering the intensity and duration of exposure in the risk model being examined. Establishing a dose–response relationship for an exposure supports a causal association, and the slope of the exposure–response relationship allows an estimation of the reduction in disease burden that might occur with a reduction in the workplace exposure. Both of these issues are important in establishing safe levels of exposure in the working environment.

Estimation of the reduction in disease burden due to an occupational exposure with the lowering of exposure levels has three components: How much disease will be prevented in those workers who begin their work exposure at the new levels? How much disease will be prevented by reducing the exposure of workers previously exposed to higher levels to these levels? and How much disease can be prevented by altering the smoking habits of the exposed workers? For those exposures for which synergism between smoking and an occupational exposure exists, the sum of these three estimates may exceed the total amount of disease that occurs in the population (Samet and Lerchen 1984; Doll and Peto 1981). If a group of asbestos workers have a fiftyfold increased risk with combined exposure and a fivefold risk with exposure only to asbestos and a tenfold risk with exposure only to cigarettes, then elimination of smoking would eliminate 90 percent of the risk (from 50 to 5) and elimination of asbestos would eliminate 80 percent of the risk (from 50 to 10). The sum of these reductions is greater than 100 percent, and points out that for prevention efforts, the synergistic effect works to potentiate the effect of the intervention.

Confounding of Occupational Exposures by Smoking Behavior

By the nature of the employing industries, most occupational exposures occur to a limited number of individuals who are often geographically clustered and who are not representative of the U.S. population. Prospective studies of cancer rates in populations that are representative of the U.S. population generally contain too few individuals with specific occupational exposures to allow analysis by occupational exposure. Therefore, most studies of occupational exposures involve populations selected on the basis of a specific exposure. Then either these selected populations of exposed workers are compared with a control group or individuals with high dose exposures are compared with individuals with low dose exposures. Validity depends upon the comparability of the groups being examined for variables that may influence cancer risk other than occupational exposure. Age is one such variable, as rates of most cancers increase with increasing age. For those cancers linked to smoking, the comparability of the smoking habits of the various exposed subjects is a second such variable. This variation may potentially confound an association between an occupational exposure and a cancer known to be associated with smoking, and control for this potential confounding may be critical for an unbiased evaluation of such an association.

Sources of Confounding

Confounding is the distortion of the apparent effect of an exposure on risk brought about by the association with other factors that can influence the outcome (Last 1983). Cigarette smoking can be a confounding factor in occupational studies through an association (either positive or negative) with the exposure in question. As described earlier in this chapter, the major determinants of smoking-related risk in a population include smoking prevalence, intensity of exposure, and duration of exposure. Each of these measures can potentially confound an occupational exposure.

Smoking Status

In occupational studies, cancer mortality in the occupational group is often compared with that in the entire population of a given geographic area. Age-specific death rates are available for the U.S. population on an annual basis and can be used to develop an age- and calendar-year-adjusted overall expected number of deaths, or a cause-specific expected number of deaths, for the population of workers being examined. The ratio of the actual number of deaths in the exposed population compared with the expected number in the general population, multiplied by 100, is referred to as a standardized mortality ratio (SMR) for the exposed population. The SMR may

be based on national mortality data or on data from the geographic location of the exposure group. In addition to providing a control population, the use of SMRs also adjusts for differences in age distribution between the exposed population and the population on which the SMR is based.

Cigarette smoking behavior is not uniformly distributed throughout the U.S. population. As demonstrated in the preceding chapter, there are substantial differences in smoking behavior among men and women, blacks and whites, different age groups, and different occupations. It is not surprising, therefore, that the smoking behavior of selected populations of exposed workers might differ markedly from the average for the U.S. population, and these differences would be expected to influence the SMR for smoking-related cancers.

Axelson (1978) has suggested that the effect on the SMR of differences in smoking habits could be estimated by dividing the population being examined into various smoking categories, multiplying the proportion of the population in that smoking category by the relative risk of developing disease produced by that smoking category, and summing the resultant numbers. The ratio of this number, calculated for the exposed population and compared with the number for the population on which the SMR is based, is then a multiplier that can be used to evaluate the effect on the SMR of the smoking habits of the exposed population.

In its simplest form this calculation would use only the proportion of smokers and nonsmokers in the population and a single relative risk number for the smokers. The effect that differences in smoking habits might have on the SMR for three different relative risks due to smoking is shown in Table 2. These different relative risks correspond approximately to the different relative risks for different sites of cancer associated with smoking (US DHHS 1982). Blair and colleagues (1985) have compared the crude and smoking-adjusted SMRs for different job categories in the population of the U.S. veterans study. They used four categories: smoker, never smoked, ex-smoker, and other. In general, adjustment for smoking did not substantially alter the SMRs for lung cancer (R 0.88), and the differences were small for most job categories (the largest difference between crude and adjusted SMR, 68.0).

Measures of Smoking Intensity

The risks due to smoking increase with increasing number of cigarettes smoked per day and depth of inhalation (Table 1) (US DHHS 1982). A calculation, similar to the one in the preceding section, can be performed using separate risk estimates for light smokers and heavy smokers and for ex-smokers. The magnitude of the effect on the SMR for lung cancer of a range of different smoking

TABLE 2.—Effect of differences in smoking prevalence on the relative risk of an occupational group compared with a control group

Assumed risk due to smoking	Proportion of smokers in control group	Proportion of smokers in exposed group				
		.1	.3	.5	.7	.9
2	.1	1.00	1.18	1.36	1.55	1.73
	.3	.85	1.00	1.15	1.31	1.46
	.5	.73	.87	1.00	1.13	1.27
	.7	.65	.76	.88	1.00	1.12
	.9	.58	.68	.79	.89	1.00
5	.1	1.00	1.57	2.14	2.71	3.29
	.3	.64	1.00	1.36	1.73	2.09
	.5	.47	.73	1.00	1.27	1.53
	.7	.37	.58	.79	1.00	1.21
	.9	.30	.48	.65	.83	1.00
10	.1	1.00	1.95	2.89	3.84	4.79
	.3	.51	1.00	1.49	1.97	2.46
	.5	.35	.67	1.00	1.33	1.65
	.7	.26	.51	.75	1.00	1.25
	.9	.21	.41	.60	.80	1.00

prevalences and dosages is shown in Table 3, calculated using a relative risk of 7 for smokers of less than one pack per day, 20 for smokers of over one pack per day, and 4 for ex-smokers. These relative risks were drawn from the major prospective mortality studies on smoking (US DHHS 1982). The proportions of smokers and ex-smokers in the population and the percentage of smokers who smoke more than 20 cigarettes per day were drawn from the data presented in the preceding chapter for the U.S. population between the ages of 20 and 64. On the basis of the data, the current differences in smoking patterns between blue-collar men and the total male population might be expected to result in a 10.2 percent elevation in the SMR for lung cancer. A hypothetical population with a prevalence of current smoking of 80 percent might have a 59.9 percent increase in the lung cancer SMR. Correspondingly, a population with a low smoking prevalence might have a 45.1 percent reduction in the SMR. These numbers are similar to those calculated for the Swedish population by Axelson (1978) as outer limits of the adjustment that might need to be made in lung cancer SMRs, secondary to differences in smoking patterns in an occupationally exposed population.

One of the basic assumptions made in the risk adjustment calculations described is that differences in smoking behavior (and the resultant risk) can be described by simple prevalence numbers (percentage of smokers, never smokers, and ex-smokers) or by using

TABLE 3.—Effect of differences in smoking prevalence on the standardized mortality ratio for lung cancer

Group	Smoking status					SMR multiplier
	Total	Current		Former	Never	
		< 20	≥ 20			
U.S. population	40.9	29.4	70.6	40.0	19.1	1.0
White collar	39.9	27.8	72.2	40.8	19.7	0.994
Blue collar	47.1	28.2	71.8	34.8	18.1	1.102
Hypothetical low	20.0	29.4	70.6	20.0	60.0	0.549
Hypothetical high	80.0	29.4	70.6	10.0	10.0	1.599

a division of current smoking prevalence into heavy smokers or light smokers. Other characteristics of smoking behavior have also been shown to influence lung cancer risk, including depth of inhalation, age of initiation (duration), and tar and nicotine yield of the cigarette smoked (US DHHS 1981, 1982). The differences in lung cancer relative risks among male smokers in the ACS study of 1 million men and women resulting from differences in depth of inhalation and age of initiation are presented in Table 1. It is apparent that substantial differences in lung cancer mortality ratios (up to fivefold) can occur within the broad category of smokers because of differences in the various dosage measures. It also appears that, in general, the difference in mortality ratios between the highest and lowest exposure categories was greater in the older age group than in the younger age group.

When the SMR is based on the general population, in which smoking behavior is in the middle range of the dosage measures in Table 1, it is unlikely that differences in behaviors between an exposed population and the general population would equal the differences between the highest and lowest dosage categories. However, sizable differences may occur, and the values shown in Table 1 can be used to estimate the impact of these differences. If the lowest age of initiation (under 15 years) were used as the risk for the exposed population, and the risk for an age of initiation of age 20 to 24 were used for the control population, there would be a 30 percent increase (using one risk value for all current smokers) in the SMRs listed in Table 3. This would increase the SMR for the hypothetical high smoking prevalence population to 207.4. A corresponding adjustment for a difference in depth of inhalation could increase

hese numbers even further. However, because there is almost ertainly some correlation among the various dosage measures smokers of higher numbers of cigarettes per day are more likely to nhale and to have begun smoking at an earlier age), it is not valid to reat these numbers as independent measures of risk. It does seem lear, however, that substantial variations can occur in the "expect-ed SMR" for a population, based on differences in smoking preva-ence, differences in number of cigarettes smoked per day, and probably differences in age of initiation. These adjustments suggest that SMRs in excess of 200 may occur owing to differences in smoking patterns and differences in depth of inhalation. The use of high tar and nicotine cigarettes might increase the SMR even further.

In the description of differences in smoking patterns by occupation presented in the preceding chapter, only modest differences between blue-collar workers and white-collar workers were found for age of initiation and number of cigarettes smoked per day. However, larger differences in these dosage measures are present among some of the subcategories of blue-collar and white-collar workers. Substantial variation from national norms in the various dosage measures may also occur because of sampling and selection bias in the small population samples that are often a real limitation in occupational studies. Even in larger studies, such as the study of 17,800 asbestos insulation workers (Hammond et al. 1979), substantial differences between the asbestos-exposed workers and the general population in number of cigarettes smoked per day are demonstrable (82.8 percent of the asbestos workers smoked more than 20 cigarettes per day in contrast with 68.5 percent of the men in the general population).

Failure to control for differences in smoking behavior may lead to a spurious impression of interaction. A spurious interaction produced by differences in smoke dose has a greater public health significance when the outcome is an apparent antagonism rather than a synergism. If the workers who smoke and are exposed to a given agent smoke fewer cigarettes per day, or began smoking later in life than the control population, an apparent protective effect (i.e., a less than additive effect) of the occupational exposure may result. In this setting, if the population of nonsmokers is too small to evaluate the effects of the occupational agent, only the biased estimate of the agent's effect on smokers will be available; the spurious antagonism may mask the effect of an occupational carcinogen by lowering the rate of lung cancer in the workers with combined exposure. A lower number of cigarettes smoked per day may be a relatively frequent confounder in worksites where smoking is not allowed during working hours, and a later age of initiation may exist in workforces with higher education levels. Thus, lack of information on smoking may lead to biased estimates of the effect of

an occupational agent, and even to the impression that the agent has no effect. This potential for missing the effects of an occupational carcinogen makes the incorporation of dosage data a critical part of the consideration of statistical interactions.

This discussion has used examples in which differences in smoking dosage measures resulted in spurious interactions between smoking and occupational exposures. However, the same potential exists for differences in occupational exposure dose between smokers and nonsmokers in the exposed population. If the smokers in the exposed population have a greater exposure to an occupational carcinogen than the nonsmokers, then the effect of combined exposure might be expected to appear to be greater than additive.

A companion question of "dosage" measurement among the smokers in occupational studies is how to classify pipe and cigar smokers and former smokers. Pipe and cigar smokers have a lower risk of developing lung cancer (but not oral cancer) than cigarette smokers and are distributed differently by age, reflecting the greater use of pipes and cigars by older men (US DHEW 1979). To the extent that differences in the use of pipes and cigars exist among exposed groups and control populations, the effects of smoking may be confounded if pipe and cigar smokers are classified in the study as smokers. Pipe and cigar smokers should be either analyzed as a separate category, or if the number of subjects is too small for separate analysis, they may be combined with light smokers as part of a dose–response relationship. A similar problem arises with former smokers. The lung cancer risk in former smokers declines with the increasing duration of cessation. Few people begin to smoke after age 25, and the percentage of the population who have quit smoking increases with increasing age. Many occupational settings have been the focus of intensive cessation efforts, particularly those worksites where an increased lung cancer risk has been established or suspected. These efforts, as well as the other previously described reasons for differences in smoking patterns, may make the prevalence and age distribution of former smokers in an occupationally exposed population different from that in a control population; therefore, former smokers should not be included with current smokers in an analysis of occupational exposures but should be treated as a separate category.

One of the methods that has been used to control for the differences in smoking between control groups and exposed populations, or between cases and controls (Liddell et al. 1984), is to examine the dose–response relationships of smoking and occupational exposure for lung cancer. An example of such an analysis performed on a group of asbestos miners using a case–control approach is presented in Table 4. The risk of developing lung cancer is shown to increase with increasing cumulative asbestos exposure in

TABLE 4.—Risks of lung cancer, by cigarette smoking and asbestos exposure, relative to all 223 cases and 715 referents for whom smoking histories were reliable; unmatched analysis

Pack-years[1]		Exposure accumulated up to 9 years before death of case (f/mL)•y			
		Low (<100)	Medium (<1,000)	High and very high (≥1,000)	All
0	Number of cases	6	7	10	23
	Number of referents	103	61	37	201
	Relative risk	0.19	0.37	0.87	0.37
1, <40	Number of cases	29	27	34	90
	Number of referents	123	93	63	279
	Relative risk	0.76	0.93	1.73	1.03
≥40	Number of cases	40	35	35	110
	Number of referents	117	79	39	235
	Relative risk	1.10	1.42	2.88	1.50
All	Number of cases	75	69	79	223
	Number of referents	343	233	139	715
	Relative risk	0.70	0.95	1.82	1.00

[1] Number of cigarettes a day/20 x duration in years.
SOURCE: Liddell et al. (1984).

all three categories of smoking dose. Stratification is useful for examining exposure–response relationships, an important element in establishing a causal association between a given exposure and lung cancer.

If stratification is used to control the confounding between smoking and an occupational exposure, careful consideration must be given to the relative magnitudes of the effects of smoking and occupational exposure on lung cancer risks when determining the number of smoking dose categories compared with the number of occupational exposure dose categories. As discussed elsewhere in this Report, the prevalence of smoking has been higher among men born between 1910 and 1930 than among men born in later decades. This cohort of men represents the older workers in many occupationally exposed populations, and it is these same workers who were previously exposed to levels of occupational agents that substantially exceeded the levels currently experienced. Thus, populations of older workers have had higher cumulative exposures to occupational agents than their younger peers at the same age, and have also had higher cumulative exposure to cigarette smoke than their younger peers at the same age. The result may be a residual confounding between cumulative occupational exposure and cumulative smoke exposure in assessing the effects of these two exposures. If the

magnitude of the effect of smoking is large compared with the magnitude of the effect of the occupational exposure, and few broad categories of smoking status are used with a greater number of categories of occupational exposure, then higher levels of smoking dose may occur with increasing occupational exposure dose category, generating a spurious dose–response relationship. Correspondingly, too few occupational exposure categories may result in a spurious strengthening of the dose–response relationship present for smoking. The total number of categories that can be used in this kind of analysis is usually limited by the number of lung cancer patients available for analysis; therefore, the distribution of the dosage categories to smoking and to the occupational exposure should reflect the relative magnitude of the effects of the separate exposures on lung cancer risk.

Duration of Exposure

In models of lung cancer risk due to smoking behavior, separate terms for intensity of smoking and duration are commonly included. In a risk model developed by Doll and Peto (1978) for the study of British physicians, the term for intensity of exposure was raised to the second power and the term for duration of exposure was raised to the power of 4.5.

Confounding may arise because of correlation between age and duration of exposure. Because of the importance of duration of exposure (and its covariate age) on lung cancer risk, the majority of the lung cancer cases will develop in the older members of a population. Correspondingly it is the smoking prevalence and dosage among these older workers that will largely determine the lung cancer risk for the population. The mean prevalence or mean dosage measures for the population do not take into account the effect of duration of exposure on the lung cancer risk. In a comparison of populations with different age distributions of smoking prevalence, or of the prevalence of heavy smokers, the population with the higher prevalence in the older age ranges will have the higher risk.

A final source of concern in examining the relationship between occupational exposure and lung cancer in cigarette smokers is generated by the lag time between the exposure to a carcinogen and the clinical manifestation of lung cancer. This lag time is a combination of the induction period (the time from exposure to disease initiation) and the latent period (the time from disease initiation to clinical manifestation) (Rothman 1981). This lag period is not fixed, but rather has a broad distribution over perhaps 50 or more years (Nicholson et al. 1982).

Epidemiologically, the shortest lag times are identified by the interval between the age of onset of exposure and the age when an increased relative risk can first be demonstrated secondary to the

exposure. For some exposures, once the exposure period has exceeded the shortest lag time, the relative risk often increases rapidly with increasing duration of exposure (Nicholson et al. 1982), resulting in a dramatic increase in disease rates with increasing age. It appears that the shortest lag period for smoking-induced lung cancers is in the range of 15 to 20 years, as demonstrated by the rise in lung cancer death rates that begins after age 30 to 35. The lag period for occupational carcinogens in lung cancer is not well characterized, but some agents have lag times similar to that found with smoking (Nicholson et al. 1982; Selikoff and Lee 1978). However, the onset of exposure to cigarettes and occupational carcinogens may occur at substantially different ages. Any such difference needs to be considered when examining the interactions of occupational exposures and smoking.

Ideally, the study of an occupationally exposed cohort would follow the entire cohort until the last survivor had died, so that late effects of exposures would not be missed. The reality of examining working populations and the need for timely assessment of existing risks makes the examination of workers at a variety of ages the norm in epidemiologic studies. In this setting, careful consideration of the differences in age of onset of smoking and of occupational exposures is necessary if the effects of occupational exposure are not to be missed or underestimated. For example, assuming that the average age of onset of smoking is 15 and the average age of onset of a particular occupational exposure is 25, the combined exposure effect is one of equal and additive risks of lung cancer and the lag time for both agents is 20 years. The lung cancer risk due to smoking would begin to increase at age 35, but because of the 10-year difference in age of onset of exposure, the risk due to the occupational exposure would not begin to be expressed until age 45, and even then would appear to be much smaller than the risk due to smoking because of the effects of the longer duration of exposure to cigarettes. If the cohort of workers with these two exposures is relatively young, with few older workers, then the effect of an occupational exposure may be missed or substantially underestimated. A similar concern exists when examining an agent that was introduced into the workplace 20 to 30 years ago. The cohort of exposed workers would represent a cross-section of ages, and therefore a cross-section of smoking habit durations. An additive risk effect of the occupational exposure would be small in comparison with the cumulative risk secondary to smoking in the older workers, and the number of cases of lung cancer in young workers (where the risk effects might be more equal) would be small. Again, the effect of an occupational carcinogen could easily be missed in this setting.

This discussion uses a simple statistical model of independent additive effects in concert with a biological concept of lag time.

Interpretation based on this kind of biologic extrapolation of statistical concepts is hazardous at best; nevertheless, some consideration of the differences in the age of onset of exposure should be part of both the biologic and the statistical considerations of the interactions between smoking and occupational exposures.

Control of Confounding

The examination of the risk associated with an occupational exposure generally requires a comparison group. Prospective mortality studies of the general population generally have too few individuals with the exposures of interest to allow analysis. Therefore, cohort and case–control formats have commonly been used. The control groups in either of these formats may be external (i.e., separate population) or internal (i.e., workers with high exposure compared with workers with lower exposure). A variety of methods have been used to deal with the confounding of occupational exposure by cigarette smoking.

Comparisons Using External Control Populations

Common external control populations are the national or regional populations. Death rates in these populations can be used to generate age- and time-adjusted expected numbers of deaths for the exposed population, with the ratio of actual deaths to expected deaths as the SMR. The large numbers of deaths in these large control populations results in relatively stable death rates over time for the common causes of death, and the smoking habits of these populations are often available from national or regional survey data. However, the smoking habits of the population are not known in relation to the cause of death, which limits the use of this data to control the confounding of occupational exposure by smoking in occupational cohorts. If the smoking habits of the workforce are also known, then the magnitude of the effect that the differences in smoking habits might have on the SMR can be estimated by assigning risk values to the proportions of the populations in different smoking categories (as described in the section on sources of confounding) (Axelson 1978). This adjustment for differences in smoking prevalence ignores trends over time as well as a variety of other potential sources of confounding. However, when this approach is used, the smoking-adjusted SMR alters the expected value of the SMR from the value of 100 that was expected prior to adjustment for smoking.

An alternative approach is to use an external control population for whom the smoking habits are known in relation to the causes of death. The use of a control population with known smoking habits allows the direct comparison of populations of smokers and non-

smokers with and without the exposure being investigated. These direct comparisons allow an examination of the risk of the occupational exposure in the absence of smoking (i.e., in never smokers) and also the examination of potential interactions between smoking and occupational exposures. A study may be constructed to prospectively or retrospectively examine the lung cancer death rates in a cohort of occupationally exposed workers compared with a control population, or a group of patients with lung cancer may be identified and matched with a set of controls without lung cancer in order to examine the frequency of a given occupational exposure in the two groups. In examining lung cancer risk, it is important that the control population be similar to the exposed population in age, ethnicity, socioeconomic status, and geographic location.

In general, studies are designed to be able to identify levels of lung cancer risk due to occupational carcinogens that are lower than the level of risk due to smoking. This potential difference in magnitude of effect needs to be assessed carefully when considering the level of detail with which the smoking data are obtained and examined.

The selection of a control group for an occupational study is often influenced by the ease with which data can be collected as well as by the comparability of the control group with the exposed workers. Control groups can be selected from unexposed workers in the same plant, from workers in different plants where no exposure occurred, from populations selected from the same geographic locations as the workers, and from populations being followed as part of other epidemiologic investigations. Some of these control groups may have substantial differences in smoking behavior from the exposed group. For example, if management and administrative employees are included in the control group, the prevalence of smoking in the control population or in comparison with a blue-collar exposed group may be reduced. Similarly, controls selected from different worksites may have different smoking patterns owing to differences in work rules, age of employees, or other demographic factors, or simply by chance. Populations drawn from other epidemiologic studies may also have different smoking patterns, and the mode of determination and definition of smoking status may be different from that used in the exposed group.

A common method of controlling for the confounding due to smoking is to separately examine smokers, nonsmokers, and former smokers. This allows examination of the independent effects as well as of the interactions; however, the examination of smoking patterns represents slightly different challenges in each of these groups.

Lung cancer risks may be examined in nonsmoking populations of occupationally exposed and nonexposed individuals for two separate reasons. First, such analyses can establish whether a risk due to occupational exposure occurs in the absence of cigarette smoking or

whether exposure only modifies the effect of smoking. Second, nonsmokers represent the lowest dosage category in examining the dose–response relationship for smoking. The demonstration of an effect of an occupational exposure in the absence of cigarette smoking requires a population of lifelong nonsmokers who have neither smoked cigarettes or cigars or used a pipe. In contrast, when a dose–response relationship is being examined, it would not be unreasonable to combine never smokers with pipe and cigar smokers, or even with light smokers, as a low dose group for lung cancer risk (pipe and cigar smokers should not be included in the low dose group for oral cancer risk). For exposures with modest increases in lung cancer risk, the low prevalence of never smoking status, coupled with the low expected risk of lung cancer in this group, means that large populations of workers must be examined in order to define the risk of exposure in the absence of smoking. Most occupational studies are limited by the size of the workforce being examined, and therefore, it is often necessary to combine never smokers with low smoking risk groups in order to have an adequate sample size. Once this combination has taken place, the study can examine only the effect of low smoke exposure coupled with occupational exposure, rather than the effects of occupational exposure in the absence of smoke exposure.

The low prevalence in many current workforces of people who have never smoked and the low risk of lung cancer in this group generally means that only a very few lung cancer deaths occur in this group, limiting the number of deaths for which to perform an analysis of the effects of an occupational exposure in the absence of smoking. For example, in the large study of asbestos insulation workers (Hammond et al. 1979), only 5 lung cancer deaths were recorded in nonsmokers out of more than 8,000 asbestos-exposed workers (smokers and nonsmokers included) whose smoking habits were known. Drawing inferences from small numbers of lung cancer cases is necessary in occupational studies, but two important caveats should be considered. First, it is essential that lung cancer patients placed in the never smoking category are actually individuals who have never smoked. The inclusion of even modest numbers of misclassified smokers or light smokers may increase the number of lung cancers over that expected on the basis of the risks in the never smoker, nonexposed control population. For this reason it is critical that the data on smoking habits be accurate and obtained in the same way in the exposed population as in the control population. When the level of monetary compensation for occupational disability may be influenced by smoking status, workers may be motivated to define themselves as never having smoked, regardless of their actual smoking status. In many studies the determination of smoking status is made for the living subjects by questionnaire or interview