Complex Mixtures and Indoor Air Pollution: Overview of Epidemiologic Methods
Noel S. Weiss

Department of Epidemiology, School of Public Health and Community Medicine, University of Washington, Seattle, WA 98195, and Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98104

The likelihood of an epidemiologic study correctly identifying an adverse health outcome associated with exposure to indoor air pollutants is increased if a) substantial variation exists in the frequency or level of exposure among study subjects otherwise at similar risk of the health outcome; b) the number of study subjects or study communities is large; c) the health outcome can be assessed with accuracy; d) relevant exposure levels can be measured with accuracy; e) an unbiased sample of exposed and nonexposed subjects is selected for study; and f) other determinants of the adverse health outcome can be measured. Nonetheless, given a strong enough impact of exposure to one pollutant or a mixture of pollutants on the risk of illness, it is possible for epidemiologic studies to discern a relation even if only some of the above circumstances are present. — Environ Health Perspect 101(Suppl 4):179-181 (1993).

Key Words: Air pollution, epidemiology

Epidemiology can be thought of as the study of the variation in disease occurrence and of the reasons for that variation. Operationally, it involves making observations in individuals or groups of individuals on the rates of disease associated with different levels of an exposure or characteristic, followed by inferences concerning the basis for any differences in rates seen. At its simplest, epidemiology can involve nothing more than seeking to correlate published rates of illness in various population groups with levels of present or past exposure in such groups. However, it generally is true that stronger inferences can be based on studies of the occurrence of illness in exposed and nonexposed individuals. Such studies occasionally involve randomization of individuals to differing environmental exposures to determine if the subsequent rate of illness (or marker of illness) differs among exposure groups. More commonly, no randomization is done, but investigators simply observe rates of illness in persons who happen to have differing levels of exposure (cohort studies). Also, especially for health outcomes that are uncommon, it is possible to identify persons with and without a disease and attempt to retrospectively reconstitute exposures that persons in each group had sustained (case-control studies).

For an epidemiologic study to provide useful information regarding causes of the disease, several circumstances need to be met. These circumstances are discussed below:

A. Among individuals otherwise at similar risk of the disease, there exists substantial variation in the frequency or level of exposure. From an epidemiologist’s point of view, this circumstance is best met when the variation occurs within members of a community (e.g., the presence of both cigarette smokers and nonsmokers in a given population). However, when dealing with certain exposures (e.g., outdoor pollution from acid aerosols and oxidants or from arsenic), the exposure may be communitywide, with little variation among individuals within that community. In this situation, it becomes particularly necessary to make comparisons among populations of differing exposure status (e.g., air pollution levels) rather than among individuals within the same population. In many instances, comparisons among populations are facilitated by the fact that routine data are available for a variety of health outcomes (e.g., mortality, cancer incidence) on a large number of populations over a long period of time. Nonetheless, studies that compare populations rarely can be used for anything but the generation of hypotheses regarding disease etiology, because a substantial degree of movement of individuals between communities occurs in most parts of the world in which these studies are likely to be conducted. This would generally be expected to dilute any true association between communitywide exposures and disease occurrence. Also, other bases for a difference in rates among populations are often quite hard to measure and therefore cannot be taken into account when looking at the exposure of interest.

For these reasons, some investigators have attempted to study communitywide exposures on health outcomes by returning to the study of individual persons within a community, exploiting the substantial degree of migration that would have occurred in years past. For example, in a study of cancer in relation to ingestion of asbestos in drinking water, Polissar et al. (1) compared persons with and without cancer who resided in one western Washington county. These cases and controls were contrasted with respect to the amount of time they had lived in those particular areas of western Washington in which there had been an extraordinarily high concentration of asbestos in the water supply. Clearly, this approach can be successful only if the induction period of the disease from the exposure in question is reasonably long.

For residential exposures that truly do vary within a community, this tendency of persons to change households frequently will act to minimize variation among individuals in that community. For example, Lubin et al. (2) note that among Americans in the 1980s there had been a change of household on the average of every 5 years. If one were attempting to study cancer in relation to household radon exposure, for example, movement between households of differing radon levels would tend to neutralize the more extreme differences that might be present if individuals had resided in a single household for a longer period of time.

Occasionally, there will not only be interindividual quantitative differences in exposure (e.g., levels of intensity or duration...
of exposure) but qualitative differences as well. Studies of individuals (or groups of individuals) who vary with regard to type of exposure can suggest what aspect of exposure might be important in disease etiology. For example, the observation that occupational exposure to amphibole, more than chrysotile, asbestos is associated with a particularly high risk of mesothelioma and lung cancer (3) has

(a) provided hypotheses regarding the pathogenesis of asbestos carcinogenicity and (b) served, in some countries, as the basis for different standards for permissible workplace air levels of amphibole versus chrysotile asbestos.

B. Whether of individuals or communities, the number of units being compared need be large enough to reliably identify an adverse health effect of the exposure if one is present. If one or more indoor air pollutants have a substantial relative impact on the occurrence of a disease, it generally is possible to identify this in a study of but modest size. For example, once mesothelioma was identified as such, a study of only a small number of individuals with and without this condition was needed to determine that inhalation of asbestos fibers was associated strongly with its occurrence. However, for many indoor air pollutants, there are reasons to believe that the true impact on disease occurrence, if any, would be small in magnitude given the relatively low levels of exposure to these pollutants and the limited variation in exposure to them in members of the population. The detection of small relative increases in disease incidence can require a study that includes a very large number of subjects, even if exposure status can be measured accurately and possible confounding factors can be taken into account. Some strategies for achieving a large number of subjects have included combining in a single study exposed groups that are scattered over a wide geographic range. For example, in attempting to evaluate the influence of occupational inhalation of formaldehyde on the occurrence of lung and other forms of cancer, individuals exposed to formaldehyde in a number of different work settings and industrial processes in a variety of locations in the United States were enrolled in a collaborative study (4). By means of meta-analysis (5), one can formally aggregate the results of multiple studies that pertain to the health impact of a particular exposure.

C. The health outcome can be assessed with accuracy and in an unbiased way. Obviously, the inability to recognize distinctive pathologic process as such will impair our ability to recognize the determinants of that process. It was not until the last half of this century that mesothelioma was identified regularly as being present in patients who truly had this malignancy. Had mesothelioma been routinely diagnosed earlier in years, undoubtedly our understanding of the carcinogenic potential of asbestos fiber inhalation would have been achieved earlier as well.

Inaccurate assessment of health outcomes also can give rise to false positive associations with respiratory exposures. This is particularly true when the outcome is defined solely on the basis of symptoms. When knowledge of a person's exposure status could influence his or her reporting of these symptoms, great care has to be taken to standardize assessment between exposed and unexposed subjects. Occasionally, it will be necessary to focus the analysis on the occurrence of symptoms of great severity. For example, in their study of possible neurologic sequelae of swine flu vaccination, Marks and Halpin (6) labeled only patients with bilateral lower motor neuron weakness of acute onset as having Guillain-Barré syndrome. They feared that, because of the concern that many patients and their physicians had regarding this vaccine, less specific neurologic illnesses would be identified more completely in vaccinated than in unvaccinated persons.

D. Exposure levels can be (or have been) measured accurately and at the appropriate time relative to the induction period of the disease under study. In many studies, whether cohort or case–control in type, the cases of disease have occurred already by the time of the study. Exposures that have occurred earlier in time need to be assessed. One way of doing this is to ask subjects, both those with and without disease, about their prior exposures. An advantage of this approach is that information can be sought about several different time periods. The primary disadvantage of the approach, however, is the relative imprecision with which the information generally can be provided. While persons might know they have been exposed to some extent to environmental tobacco smoke, for example, they would find it difficult to quantify this exposure in an accurate way. For other types of exposure (e.g., radon), no subjective assessment is possible. Direct measurements of present exposures can be made, but responsibility falls on the investigator to take steps to assess their comparability to exposures that the subject sustained in the past. For some (e.g., residential radon), this is more feasible than for others, because prior radon exposures can be estimated from present ones given the known decay of this element combined with additional information on structural and other alterations to the residence.

At first glance, it would seem that studies in which measurements are made at the time the study begins, with subsequent monitoring of the occurrence of illness, would have substantial advantages over those that try to ascertain exposures in a retrospective way. However, there are at least two important limitations of these prospective studies: (a) Unless the follow-up period is very long, the study population very large, or the disease under study very common, the number of health outcomes that occur may be small and may yield highly tentative results. (b) Depending on the length of the induction period for the disease, single measurements made at the start of the study may not be relevant for long to disease occurrence. For example, in their prospective study of environmental tobacco smoke in relation to the occurrence of fatal coronary heart disease, Garland et al. (7) assessed exposure to spouse’s smoking via an interview. Among members of this cohort, the occurrence of fatal heart disease was then monitored during the next decade but with no additional information regarding continued exposure to spouse’s smoking. If exposure to environmental tobacco smoke predisposes to the occurrence of fatal heart disease through a relatively short-term mechanism (perhaps via acute toxicity of elevated levels of carboxyhemoglobin), this research approach would be a relatively insensitive means of addressing the hypothesis, given the occurrence of changes in the exposure to spouse’s smoking during the extended follow-up.

E. An unbiased sample of exposed and nonexposed individuals has been selected for study. While this is a concern in any study, it is a particular problem for those that are cross-sectional in nature. In such a study, exposed and nonexposed individuals are contrasted for their prevalence of disease. A seriously biased underestimate of the health impact of the exposure will be obtained if persons who have suffered disease because of the exposure are no longer present at the time of sampling (e.g., through premature retirement from a hazardous occupation or due to death). For example, in the 1940s, Fleischer et al. (8) noted only a low prevalence of asbestososis among men who had been employed as pipe coverers in a shipyard and who, through this employment, had been exposed to asbestos. Undoubtedly, the selective removal from employment of those who already had been affected by asbestos led to
the overly optimistic conclusion by the authors that there was little to be feared in terms of levels of asbestos exposure present in that occupation at that time.

F. Other factors besides the exposure in question that relate to the occurrence of disease have been (or can be) measured as well. Measurement of such factors will enable, first, the control of potential confounding effects of these other variables (and thus the prevention of the distortion of the true association between the exposure and disease). For example, in a study of respiratory infection during childhood in relation to exposure to environmental tobacco smoke and nitrogen dioxide, it would be important to ascertain such things as exposure to infected individuals, household crowding, etc. Second, the characterization of other exposures can enhance the power of the analysis by allowing an examination of the effect of the exposure in question according to the presence or absence (or level of) other risk factors for disease. If, for example, domestic exposure to radon were a cause of lung cancer only in the presence of active cigarette smoking, an analysis that failed to examine the association separately in cigarette smokers and nonsmokers would provide a blurred result. On the other hand, if domestic radon exposure and cigarette smoking acted via separate causal pathways to produce the disease (as appears at least in part to be the case for occupational radon exposure and cigarette smoking) (9,10), then the relative impact of exposure to domestic radon would be far more discernible in nonsmokers with their low background rate of lung cancer than among cigarette smokers in whom there is a high background rate (11).

Conclusions
The foregoing has indicated some of the major threats to the sensitivity and validity of epidemiologic studies of the health consequences of indoor air pollution. While these threats are real, it would not be prudent to allow their specter to paralyze prospective investigators and discourage them from performing research in this area. Not all of the above criteria need be met in order for a study to produce some useful information. For example, the hypothesis that military service during the Vietnam war era predisposed people to the subsequent occurrence of suicide received strong support from a study (12) that found an increased rate of suicide among men whose birthdates made them eligible to be drafted during that time. Despite the great imprecision with which actual military service in Vietnam was assessed (it is estimated that only 25% of individuals with draft eligible birthdates even entered the armed forces) and the modest size of the association (the study observed a relative risk of 1.13), the randomized nature of the investigation and its ability to neutralize the effect of potential confounding variables made for convincing results.

Imprecise exposure assessment also was a problem in a cancer registry-based study of the hypothesis that homosexual men are at increased risk for the occurrence of anal cancer (13). Registry data do not provide information on sexual preference, but they do contain data regarding marital status. The investigators found that the percentage of men with anal cancer who had never been married was more than three times that of demographically comparable men with colon or rectal cancer. Of course, being a single male is hardly an accurate predictor of homosexual preference. Nonetheless, given the exceedingly strong association between a history of anal intercourse and anal cancer (found subsequently in response to the registry-based study), even a study that measured exposure status as imprecisely as this study was able to make a contribution.

The important findings in these last two studies, studies that had serious flaws as measured by the criteria that have been put forth here, should serve to dispel the notion that only perfect studies will permit progress toward understanding the harmful effects of indoor air pollution on health. Imperfect studies, properly interpreted, are far better than none at all.

REFERENCES
1. Polissar L, Severson RK, Boatman ES. A case–control study of asbestos in drinking water and cancer risk. Am J Epidemiol 119:456–471 (1984).
2. Lubin JH, Samet JM, Weinberg C. Design issues in epidemiologic studies of indoor exposure to Rn and risk of lung cancer. Health Phys 59:807–817 (1990).
3. Mossman T, Gee BL. Asbestos-related diseases. N Engl J Med 320:1721–1730 (1989).
4. Blair A, Stewart P, O’Berg M, Gaffey W, Walrath J, Ward J, Bales R, Kaplan S, Cubit D. Mortality among industrial workers exposed to formaldehyde. J Natl Cancer Inst 76:1071–1084 (1986).
5. Greenland S. Quantitative methods in the review of epidemiologic literature. Epidemiol Rev 9:1–30 (1987).
6. Marks JS, Halpin TJ. Guillain-Barré syndrome in recipients of A/New Jersey influenza vaccine. JAMA 243:2490–2494 (1980).
7. Garland C, Barrett-Connor E, Suarez L, Criqui MH, Wingard DL. Effects of passive smoking on ischemic heart disease mortality of non-smokers. Am J Epidemiol 121:645–650 (1985).
8. Fleisher WE, Viles FJ, Gade RL, Drinker P. A health survey of pipe covering operations in constructing naval vessels. J Ind Hyg Toxicol 28:9–16 (1946).
9. Samet JM, Kuvirt DM, Waxweiler RJ, Key CR. Uranium mining and lung cancer in Navajo men. N Engl J Med 310:1481–1484 (1984).
10. Radford EP, Renard Sr, Clair KG. Lung cancer in Swedish iron ore miners exposed to low doses of radon daughters. N Engl J Med 310:1485–1494 (1984).
11. Weiss NS, Liff JM. Accounting for the multicausal nature of disease in the design and analysis of epidemiologic studies. Am J Epidemiol 117:14–17 (1983).
12. Heart N, Newman TB, Hulley SB. Delayed effects of the military draft on mortality. N Engl J Med 314:620–624 (1986).
13. Daling JR, Weiss NS, Klopfenstein LL, Cochran LE, Chow WH, Daifuku R. Correlates of homosexual behavior and the incidence of anal cancer. JAMA 247:1988–1990 (1982).