6. Research in the present decade has established that involuntary smoking is a cause of disease, including lung cancer, in healthy nonsmokers, and that the children of parents who smoke have an increased frequency of respiratory infections and symptoms.

7. In 1964, tobacco use was considered habituating. A substantial body of evidence accumulated since then, and summarized in the 1988 Surgeon General’s Report, has established that cigarettes and other forms of tobacco are addicting. Given the prevalence of smoking, tobacco use is the Nation’s most widespread form of drug dependency.

8. Studies dating from the 1950s have consistently documented the benefits of smoking cessation for smokers in all age groups.

9. Recent evidence, including that presented in this 1989 Report of the Surgeon General, documents that cigarette smoking is a cause of cerebrovascular disease (stroke) and is associated with cancer of the uterine cervix.

Part II. The Physicochemical Nature of Tobacco

1. The estimated number of compounds in tobacco smoke exceeds 4,000, including many that are pharmacologically active, toxic, mutagenic, and carcinogenic.
2. Forty-three carcinogens have been identified in tobacco smoke.
3. Carcinogenic tobacco-specific nitrosamines are found in high concentrations in smokeless tobacco.
References


105


NATIONAL INSTITUTES OF HEALTH. NATIONAL INSTITUTE OF ENVIRONMENTAL HEALTH SCIENCES. Airborne pollutants and respiratory cancer reproductive and developmental toxicology. Environmental Health Perspectives 70, December 1986.


PETO, R., SPEIZER, F.E., COCHRANE, A.L., MOORE, F., FLETCHER, C.M., TINKER, C.M., HIGGINS, I.T.T., GRAY, R.G., RICHARDS, S.M., GILLILAND, J., NORMAN-


YAMASAKI, E., AMES, B.N. Concentration of mutagens from urine by adsorption with the nonpolar resin XAD-2: Cigarette smokers have mutagenic urine. (Detecting Carcinogens in Urine or Aqueous Liquids/Salmonella Test). Proceedings of the National Academy of Science 74(8):3555–3559, August 1977.


CHAPTER 3

CHANGES IN SMOKING-ATTRIBUTABLE MORTALITY
CONTENTS

Introduction ....................................................... 121
A Twenty-Year Perspective: 1965–85 .................................. 122
The Concept of Attributable Risk .................................... 122
Mathematics of Attributable Risk .................................... 123
Illustrative Calculation: Smoking and Lung Cancer in Women ............ 125
Uncertainties in Attributable Risk .................................... 127
Aggregation Bias Versus Statistical Precision ..................... 127
Age-Standardization ......................................... 127
Potential Biases in Applying the Results of Prospective Studies to the General Population ..................... 128
Uncertainties in Exposure .................................... 129
Errors in the Classification of Causes of Death .................... 130
Previous Estimates of Attributable Risk from Cigarette Smoking ........ 130
Populations at Risk: 1965 and 1985 ................................... 132
Cigarette Smoking and Other Forms of Tobacco Use .................... 133
Older Cohorts of Cigarette Smokers .................................. 134
Overlapping Populations at Risk ..................................... 136
Changes in the Cigarette Product .................................... 138
Other Changes in the Cigarette Smoking Population ..................... 139
Those Smokers Most at Risk in 1985 Were Also Smokers in 1965 .......... 140
Cancer Prevention Study I and Cancer Prevention Study II ................. 140
Nonsmokers’ Death Rates .......................................... 141
Current Cigarette Smokers’ Death Rates: Lung Cancer .................. 143
Current Cigarette Smokers’ Death Rates: Coronary Heart Disease ........ 144
Current Cigarette Smokers’ Death Rates: Chronic Obstructive Pulmonary Disease ............................................ 146
Estimated Relative Risks from CPS-I and CPS-II .................... 146
Endocrine and Sex-Related Cancers in Women ......................... 152
Summary ........................................................ 152
Smoking-Attributable Mortality in the United States, 1965 and 1985 ........ 153
Conclusions ....................................................... 161
References ...................................................... 162
Introduction

In 1938, Raymond Pearl reported elevated death rates among white males who smoked tobacco, especially those aged 30 to 60 years (Pearl 1938). Pearl’s study of 6,800 subjects revealed the increase in mortality risk to be highest among heavy smokers. In 1954, Hammond and Horn reported on the 20-month followup of their prospective study of 188,000 white men, aged 50 to 69 years (Hammond and Horn 1954). Death rates were highest among men who smoked cigarettes but not other tobacco products, and increased with the amount of cigarette use. Overall, the number of deaths among cigarette smokers was 57 percent greater than would be expected from non-smokers’ mortality rates. Most of the increased mortality could be attributed to deaths from cancer and especially from coronary heart disease (CHD).

In 1964, the Advisory Committee to the Surgeon General reviewed seven prospective studies of smoking and mortality, encompassing over 1.7 million entrants. For the 1.1 million male enrollees, the overall mortality ratio, defined as the observed number of deaths in current cigarette smokers divided by the number expected from nonsmokers’ rates, was 1.68. “For all seven studies,” the Committee stated, “coronary artery disease is the chief contributor to the excess number of deaths of cigarette smokers over nonsmokers, with lung cancer uniformly in second place. For all seven studies combined, coronary artery disease (with a mortality ratio of 1.7) accounts for 45 percent of the excess deaths among cigarette smokers, whereas lung cancer (with a ratio of 10.8) accounts for 16 percent” (US PHS 1964, p. 30).


The 1984 Report estimated that 80 to 90 percent of the 62,000 deaths from chronic obstructive lung disease (COLD), referred to later in this discussion as chronic obstructive pulmonary disease (COPD), in the United States in 1983 were attributable to cigarette smoking (US DHHS 1984). “Over 50,000 of the COLD deaths can therefore be considered preventable and premature since these individuals would not have died of COLD if they had not smoked” (US DHHS 1984, p. ii). In 1987, the Economic Report of the President stated, “Smoking presents the largest single source of health risk in America” (U.S. President 1987, p. 184).

This Chapter further delineates the mortality consequences of cigarette smoking in the United States. Deaths attributable to cigarette smoking are reported for two benchmark years—1965 and 1985. The Chapter focuses on the health consequences of smoking for current and former cigarette smokers. Deaths of nonsmokers caused by environmental tobacco smoke (National Research Council 1986; US DHHS 1988a) and
deaths from cigarette-related fires (Consumer Product Safety Commission 1987; Botkin 1988) are not discussed, nor are the morbidity consequences of cigarette smoking (US DHEW 1979; Rice et al. 1986).

A Twenty-Year Perspective: 1965–85

The two-decade interval, 1965–1985, was selected primarily for reasons of data availability. The year 1985 was the most recent one for which complete, nationwide, cause-specific mortality statistics were available from the National Center for Health Statistics (NCHS). Moreover, in both 1965 and 1985, questions on cigarette use were appended to the National Health Interview Survey (NHIS), a nationally representative, face-to-face interview survey that has been conducted annually by NCHS (Massey et al., 1987; NCHS 1986). In particular, 1985 was the most recent full year for which complete population-weighted data from the NHIS were available (see Chapter 5).

In addition, the years 1965 and 1985 represented the approximate midpoints of two large-scale prospective surveys of smoking and mortality among men and women in the United States, both sponsored by the American Cancer Society. In the first of these two prospective studies (Garfinkel 1980a,b, 1981; Hammond 1961, 1964a,b, 1966, 1968, 1969, 1972; Hammond and Garfinkel 1961, 1964, 1966, 1968, 1969, 1975; Hammond et al. 1976; Hammond and Seidman 1980; Lew and Garfinkel 1984, 1988), about 1 million persons were followed from 1959 through 1972. In the second study (Garfinkel 1985; Stellman and Garfinkel 1986; Stellman, Boffetta, Garfinkel 1988), about 1.2 million participants were followed from 1982 through 1988. The two studies will be referred to, respectively, as “Cancer Prevention Study I (CPS-I)” and “Cancer Prevention Study II (CPS-II).” In particular, this Chapter will present unpublished, preliminary results from the 4-year followup (1982–86) of CPS-II.

The theory, mathematics, limitations, and other methodological issues concerning the calculations of smoking-attributable mortality are described in the next section. The results of the analysis follow thereafter. Readers interested primarily in those results may proceed directly to the Section entitled “Populations at Risk: 1965 and 1985.”

The Concept of Attributable Risk

In 1953, Levin estimated that 62 to 92 percent of all male lung cancers were “attributable to cigarette smoking” (Levin 1953). Levin’s computations addressed the general problem: How many cases of a disease in a given population can be explained by the presence of a particular hazardous agent or a particular personal trait? Put differently, how many cases would have been avoided but for the presence of the agent or the trait (Doll and Peto 1981)?

In principle, the answer requires an experiment whereby disease rates are measured before and after the complete elimination of the hazardous agent or particular trait from the population of interest. Since this type of experiment is usually impractical, the most widely used approach is to estimate disease rates in representative sample populations of exposed and unexposed persons. The results are then extrapolated to the population of interest.
The phrase "cases attributable to agent A" is often used interchangeably with "cases caused by agent A." The latter term is meaningful so long as it is recognized that "caused" refers to an entire population rather than to any single, predetermined member of the population. Thus, the scientific validity of an estimate that 1,000 lives would be saved by the removal of some hazardous agent does not hinge upon naming the names of the people to be saved.

The population-based notion of causation is especially important for chronic diseases with multiple causes. Agent A, for example, may promote or enhance the disease-causing effect of agent B. A case-by-case analysis of afflicted individuals may never identify agent A as the primary cause in a single instance. Yet its elimination might substantially reduce disease incidence in the population under study.

Moreover, the concept of attributable risk generally requires a timeframe. In an assessment of the effects of removing a hazardous agent, a researcher could ask how many cases of a specific disease could be avoided in a specified time period, such as 1 year. When the disease has multiple causes, this quantity may differ from the number of cases of the disease that may eventually be avoided. By specifying a timeframe, the researcher inquires not whether such cases could be completely prevented, but whether their premature occurrence could be avoided.

For many diseases, death rates are more accessible and reliable than disease rates. Accordingly, computations of "attributable deaths" from a disease have been used in place of "attributable cases" of the disease. Because death from one cause or another is inevitable, such computations necessarily refer to a specific time period during which premature mortality may have been prevented.

Mathematics of Attributable Risk

Let $d_1$ and $d_0$, respectively, denote the incidence rates (in terms of new cases per unit time) of a particular disease among two sample cohorts—one exposed to a hazardous agent, the other unexposed. The two samples are assumed not to differ materially in any other respect, so that both would experience disease incidence $d_0$ in the absence of exposure. Accordingly, the difference $d_1 - d_0$ measures the increase in disease incidence, or absolute risk, due to the agent. Moreover, the unitless ratio $r = d_1/d_0$, termed the relative risk, measures the degree to which the hazardous exposure multiplies the baseline incidence rate. It is often employed as a measure of the epidemiologic and biological significance of an observed association between an agent and a particular disease (Lilienfeld and Lilienfeld 1980; US DHEW 1979).

In the exposed cohort, the proportion of disease cases attributable to the hazardous agent is thus equal to $\frac{s = (d_1 - d_0)d_1}{d_0}$ (which equals $(r-1)/r$). This quantity has been variously termed the assigned share or probability of causation or attributable proportion of risk among the exposed (Bond 1981; Ofstedal, Magnus, Hvinden 1968; Black and Lilienfeld 1984; National Research Council 1984; Cox 1987).

For some hazardous agents, such as cigarette smoke, the disease incidence rates $d_1$ and $d_0$ and the relative risk $r$ have been estimated directly from prospective longitudinal studies of exposed and unexposed cohorts. Alternatively, retrospective case control studies do not provide estimates of $d_1$ and $d_0$ but yield a close approximation to the rela-
tive risk $r$ when incidence of the disease is low (Cornfield 1951). Both types of studies provide estimates of the assigned share $s$.

The estimate of relative risk $r$, derived from epidemiologic studies, is then applied to the population of interest. Let $p$ denote the proportion of exposed persons in the subject population, estimated independently from survey data. Then the quantity $f = pr/[p(r-1)+1]$ is the fraction of all cases of the disease (in a given time interval) that occurs among exposed persons in the subject population. This is sometimes called the “case fraction” (Miettinen 1974). Moreover, if fraction $f$ of all cases occurs among exposed persons, and if fraction $s$ of such exposed cases is attributable to the hazardous agent, then the fraction of all cases attributable to the agent is $a = fs$. From the definitions of $f$ and $s$, the quantity $a$ can be expressed as

$$a = \frac{p(r-1)}{p(r-1) + 1}$$

This is Levin’s measure of attributable risk, also termed etiologic fraction (Miettinen 1974), attributable fraction (CDC 1987b), and population-attributable risk (MacMahon and Pugh 1970). When $a$ is expressed in percentage terms, it is often termed percent attributable risk or population-attributable risk percentage.

Equation (1) shows how the attributable risk $a$ depends upon both the relative risk $r$ and the proportion exposed $p$. Thus, an agent may be significant in the causation of disease among exposed persons so that its relative risk $r$ greatly exceeds 1. Yet that agent may cause a small proportion of all cases of the disease because exposure rates $p$ are low. Conversely, an agent that is widely prevalent (with large $p$) may contribute substantially to the total number of cases, even when its relative risk $r$ is close to unity.

As a consequence of equation (1), the logistic transformation of $a$ is

$$\log \frac{a}{1-a} = \log p + \log (r-1)$$

where log denotes the natural logarithm. Equation (2) provides a convenient method of decomposing the uncertainty in the attributable risk $a$ into two components—uncertainty in the proportion exposed $p$ and uncertainty in the relative risk $r$.

Levin’s measure of attributable risk can be generalized to cases where there are multiple levels of exposure, multiple causative agents, or confounding or stratifying variables, or when an agent can prevent a disease (Walter 1976; Miettinen 1974). In the case of multiple levels of exposure, it is convenient to let $dk$ denote the incidence rate and $rk = dk/d0$ denote the relative risk for the $k$-th exposure level. Similarly, let $pk$ denote the proportion of the subject population exposed at the $k$-th level. Then $sk = (rk-1)/rk$ is the assigned share among cases exposed at the $k$-th level. Likewise, the quantity $fk = pkrk/[\Sigma_k pk (rk-1) + 1]$, where $\Sigma_k$ denotes summation over exposure levels, is the fraction of all cases occurring among persons exposed at the $k$-th level. The generalized formula for attributable risk becomes $\Sigma_k f_k s_k$, which can be expressed as

$$a = \frac{\Sigma_k pk (rk-1)}{\Sigma_k pk (rk-1) + 1}$$

Let $D$ denote the total number of cases of disease in the population of interest in a given time interval. Then $A = aD$ is the estimated number of cases in the interval that
are attributable to the agent. The quantity \( A \) is sometimes called "attributable cases." When relative risks or exposure rates vary by age, sex, or other stratifying variables, then separate estimates of \( A \) can be made for each combination of variables.

When there are multiple causative agents, attributable risks can be computed for each agent separately and for combined exposures. Thus, if agents X and Y both have a causal role in the development of a particular disease, then the relative risk for agent X may depend upon the presence or absence of exposure to agent Y. When X and Y act synergistically, some portion of the total risk attributable to X will reflect the combined contribution of X and Y. For example, indoor exposure to radon has recently been estimated to account for about 13,300 lung cancer deaths annually in the United States (Lubin and Boice 1988). Radon exposure and cigarette smoking interact synergistically in causing lung cancer (National Research Council 1988). Of the estimated 13,000 deaths attributable to radon exposure, about 11,000 would be due to the combined effect of smoking and radon, while about 2,000 would reflect radon exposure in non-smokers (Lubin and Boice 1988).

Illustrative Calculation: Smoking and Lung Cancer in Women

Table 1 provides a detailed illustrative application of Levin's method to female deaths from lung cancer in the United States during 1985. The population of female smokers has been divided into ten exposure levels: five categories of current cigarette smokers based on the number consumed per day; and five categories of former cigarette smokers based on the length of time since quitting. For each exposure category, the upper panel shows the estimated prevalence \( p_k \), derived from the 1985 NHIS. Also given are estimates of relative risk \( r_k \) derived from the 4-year followup (1982–86) of the second American Cancer Society prospective study (Garfinkel and Stellman 1988). At each exposure level, the upper panel also shows the assigned share \( s_k \) and the case fraction \( f_k \).

The computations are summarized in the lower panel of Table 1. For both current and former smokers, as well as for all females at risk, the estimated prevalences \( p \) represent the corresponding sums \( \Sigma_k p_k \) over the prevalence rates \( p_k \) in the individual subcategories. The case fractions \( f \) likewise represent sums of individual fractions \( f_k \), while the attributable risks \( a \) are derived from the corresponding sums \( \Sigma_k s_k f_k \). Attributable deaths \( A \) are derived from the products \( aD \), where \( D = 38,687 \) lung cancer deaths among adult females in 1985.

Table 1 shows that almost two-thirds of all female lung cancer deaths occurred among women who currently smoke one pack or more daily or who have quit smoking within the last 5 years. Nine out of ten lung cancer deaths occurred in women with any history of regular cigarette use. Cigarette smoking accounted for an estimated 82 percent of lung cancer deaths in women, or 31,600 deaths in 1985. About 9,300 (or 29 percent) of the 31,600 female lung cancer deaths that were caused by smoking occurred among former smokers.

Both the prevalence rates and the relative risks in Table 1 are subject to sampling variability. By a formula analogous to equation (2), a standard error for the logistic transformation of \( a \) can be derived. Under the assumption that \( D \) has no sampling