Landmarks in the History of Cancer Epidemiology

Peter Greenwald and Barbara K. Dunn
Division of Cancer Prevention, National Cancer Institute, NIH, Bethesda, Maryland

Abstract

The application of epidemiology to cancer prevention is relatively new, although observations of the potential causes of cancer have been reported for more than 2,000 years. Cancer was generally considered incurable until the late 19th century. Only with a refined understanding of the nature of cancer and strategies for cancer treatment could a systematic approach to cancer prevention emerge. The 20th century saw the elucidation of clues to cancer causation from observed associations with population exposures to tobacco, diet, environmental chemicals, and other exogenous factors. With repeated confirmation of such associations, researchers entertained for the first time the possibility that cancer, like many of the infectious diseases of the time, might be prevented. By the mid-20th century, with antibiotics successfully addressing the majority of infectious diseases and high blood pressure treatment beginning to affect the prevalence of heart disease in a favorable direction, the focus of much of epidemiology shifted to cancer. The early emphasis was on exploring, in greater depth, the environmental, dietary, hormonal, and other exogenous exposures for their potential associations with increased cancer risk. The first major breakthrough in identifying a modifiable cancer risk factor was the documentation of an association between tobacco smoking and lung cancer. During the past four decades, epidemiologic studies have generated population data identifying risk factors for cancers at almost every body site, with many cancers having multiple risk factors. The development of technologies to identify biological molecules has facilitated the incorporation of these molecular manifestations of biological variation into epidemiologic studies, as markers of exposure as well as putative surrogate markers of cancer outcome. This technological trend has, during the past two decades, culminated in emphasis on the identification of genetic variants and their products as correlates of cancer risk, in turn, creating opportunities to incorporate the discipline of molecular/genetic epidemiology into the study of cancer prevention. Epidemiology will undoubtedly continue contributing to cancer prevention by using traditional epidemiologic study designs to address broad candidate areas of interest, with molecular/genetic epidemiology investigations honing in on promising areas to identify specific factors that can be modified with the goal of reducing risk. [Cancer Res 2009;69(6):2151–62]

Introduction

Cancer epidemiology is the study of patterns and determinants of cancer risk, outcome, and control in healthy or diseased populations. Unlike the experimental nature of most clinical research, much of epidemiology is inherently observational, the study of people with and without disease (1, 2). One of epidemiology's initial roles in supporting cancer research was the collection of descriptive data on cancer incidence and mortality to identify the health effects of cancer on society. As illustrated in Figs. 1 and 2, site-specific data on the mortality of women and men in the United States from cancer has been collected and published since 1930. Understanding mortality trends for major cancer types has focused research efforts and public policies in areas that can affect public health.

The search for associations between cancer outcomes and measurable exposures has evolved in response to increasingly sophisticated laboratory technologies to analyze "biomarkers" in tissue specimens. Molecular epidemiology is the discipline that has emerged to investigate such molecular factors. Biomarkers reflect either the body's response to specific environmental exposures ("surrogates of exposure") or innate physiologic or genetic characteristics, or they serve as putative surrogates of clinical cancer outcomes. Modern genetic epidemiology includes approaches that focus on the molecular material of genes and chromosomes in constitutional DNA, reflecting inherited cancer predisposition, as well as on tumor type–specific somatic genetic changes.

Two key features of cancer epidemiology must be kept in mind if its proper application to serving the general health and well being is to be implemented. First, caution must be exercised in inferring causality from measured associations because some factors occur in association with both a real causal factor and with the disease itself; and second, the ultimate goal is to improve the health of people, preferably via cancer prevention; observational data alone does not accomplish this. Whenever possible, data generated from epidemiologic studies, however sophisticated, must be integrated into clinical research, in which experimental manipulation of the implicated risk factors can be evaluated for its effect on cancer outcomes. In this article, we take a historical approach to surveying key areas of epidemiologic investigations into cancer risk, always with an eye to the implications of study findings in terms of public health. Our goal is not to be totally comprehensive; rather, we wish to delve into selected epidemiologic research endeavors with a critical eye grounded in a modern scientific perspective. We hope in this way to elicit the relevance of historical study findings in terms of their relevance to current and future epidemiologic investigation and especially in relation to their effect on public health.

Tobacco

The establishment of tobacco as a carcinogen is one of the signal achievements in epidemiology. Observational studies published as
early as the 1700s suggested an association of tobacco with nasal cancer among snuff workers (3), and mouth and lip cancers among pipe smokers (4). The strongest evidence supporting tobacco exposure as an etiologic factor for various cancers emerged from 20th century epidemiologic studies. In 1912, Isaac Adler published his observation that smoking tobacco may have been linked to reported increases in lung cancer cases in local hospitals (5). By the 1950s, strong epidemiologic and experimental evidence indicated that cigarette tar and other tobacco constituents produced by burning the product and inhaling the smoke were responsible for cancers of the lung (6, 7) and that an unequivocal association existed between smoking and lung cancer, especially when considering duration of smoking (6, 8). These seminal studies from the work of Ernst Wynder and Richard Doll (6, 8) eventually led to the recognition by the U.S. Surgeon General and other public health institutions and medical societies of the dangers of tobacco use in any form. This, in turn, led to important public health programs to discourage people, especially young people, from beginning smoking and to encourage those who currently smoke to quit.

From 1950 to the present, epidemiology has played a critical role in the search for the etiologic factors responsible for the dramatic increase in lung cancer incidence and mortality during the 20th century. Animal studies identified cigarette smoke constituents, including polynuclear aromatic hydrocarbons, that have carcinogenic initiator-promoter activity (9). Eventually, mechanistic studies elucidated the numerous carcinogenic genetic alterations in relevant biological pathways: xenobiotic metabolism, control of genomic instability (including DNA repair mechanisms), cell cycle checkpoints, apoptosis, telomere length, control of the microenvironment by elements such as metalloproteinases, inflammation, and growth factors (10). Complete answers to the question of cancer susceptibility would have to await the development of molecular epidemiology. This field integrates molecular biology and cancer genetics with classic epidemiologic study designs, allowing the identification of individuals or groups at increased risk or increased vulnerability to exogenous risk factors, i.e., those who may benefit from targeted cancer prevention or treatment strategies (11).

One example of the variability of risk among populations emerged from the long-held knowledge that association of \( \text{CYP1} \) enzymes with cancer risk due to tobacco exposure is due to the involvement of these enzymes in activating carcinogens such as polynuclear aromatic hydrocarbons. Specific \( \text{CYP1A1} \) polymorphisms (MspI and Ile462Val polymorphisms) were associated with increased lung cancer risk in the Japanese population, whereas only the \( \text{CYP1A1-MspI} \) polymorphism was strongly associated with risk in Western populations (10).

As early as the 1980s, the National Cancer Institute has directed antismoking efforts based on population data supplied by epidemiologists. The two largest efforts involved a clinical trial (Community Intervention Trial for Smoking Cessation) and a state level intervention program to address policy-driven initiatives to promote smoke-free environments, counter tobacco advertising and promotion, limit tobacco access and availability, and increase tobacco prices through new excise taxes (American Stop Smoking Intervention Study). The Community Intervention Trial for Smoking Cessation and American Stop Smoking Intervention Study showed that a multi-tiered approach could decrease the prevalence of smoking and increase the quit rate among current smokers (12–14).

**Diet/Body Mass Index/Physical Activity**

Although the associations between diet and disease have been known for centuries (e.g., limes to prevent scurvy in British sailors),
not until the results of rigorous epidemiologic studies were reported in the 1970s and 1980s did compelling evidence for the breadth of the association emerge (15–17). The application of epidemiologic methods to diet/cancer studies is relatively new because of the difficulty in determining the effects of specific dietary factors on risks of chronic diseases, especially at the individual level. Initially, diet/cancer epidemiology involved ecological studies of populations and suggested broad associations, such as that of red meat intake and increased colon cancer risk (18). Migrant studies provided a unique tool that, in the early days of diet/cancer epidemiologic research, focused on contrasting disease outcomes displayed by similar populations in different “ecological” settings. An early migrant study comparing the rates of gastric and colon cancers in Japanese migrating to the United States (19) indicated lower rates of gastric cancer and higher rates of colon cancer the longer immigrants lived in the United States. Patterns similar to that for colon cancer were seen for cancers of the breast, uterine corpus, and ovary among women and for prostate cancer among men.

By the mid-1990s, evidence from numerous large population studies suggested a strong association between a high intake of vegetables and fruits and reduced risks of cancers of the breast, prostate, colorectum, lung, stomach, and other sites (20). Also in the 1990s, the National Cancer Institute initiated a community-based partnership based on epidemiologic and intervention studies—the 5 A Day For Better Health Program (5 A Day)—to encourage increased vegetable and fruit intake among Americans; an evaluation in 2000 indicated a slow but steady increase in vegetable and fruit consumption among American consumers since the program’s inception (21).

Beyond overall diet, the association of cancer risk with specific dietary factors, including individual nutrients, has also been evaluated. Conflicting results have led to controversies over the roles of such dietary factors. One of the longest-running controversies involves the association of dietary fat with breast cancer: does an increase in the consumption of dietary fat increase the risk of breast cancer, and does a low-fat diet confer lower risk? A multitude of studies (case-control, cohort, meta-analyses) spanning two decades attempted to resolve the issue without reaching conclusive agreement (22, 23). The dietary fat-breast cancer controversy exemplifies the challenges of basing cancer prevention recommendations solely on observational epidemiologic studies; the need for clinical trials to test results suggested in observational studies is critical to clarify the role of lifestyle choices in cancer risk, especially in the face of conflicting epidemiologic findings.

Occasionally, epidemiologic findings seem to be inconsistent with clinical trial results, but in fact, the apparent conflict at times seems to have resulted from the failure to consider individual variability. In the case of the association between β-carotene and lung cancer, conflicting evidence in results from epidemiologic studies (foods containing β-carotene reduce lung cancer risk) and prospective clinical trials such as the Alpha-Tocopherol, Beta-Carotene (ATBC) Cancer Prevention Study, which suggested that β-carotene in supplements promotes lung cancer risk among smokers, catalyzed intense discussions to discern reasons for the inconsistency (24). A possible explanation for this apparent conflict emerged from pharmacogenomic studies conducted post-trial in the ATBC cohort. In most cases, a GSTM1 null genotype is associated with a higher risk of lung cancer when compared to a GSTM1-present genotype, regardless of treatment or years of smoking. Furthermore, the positive association between smoking duration and lung cancer is greater among GSTM1-null individuals than those with a GSTM1-positive genotype. β-carotene supplementation confers an increased risk of lung cancer at every duration of smoking with one exception: in individuals with a GSTM1-null genotype. Although all GSTM1-null individuals who fall into the longest smoking duration category have a significantly elevated risk of lung cancer compared...
to GSTM1-positives, this risk is somewhat lower in those who have had β-carotene supplementation (25). In addition to identifying a subgroup that might benefit from an intervention, this study exemplifies how clues from classic epidemiology can be obscured in the absence of in-depth characterization of individuals likely to experience a benefit (or harm). Clearly, one size doesn’t fit all. The success of dietary components as interventions must be examined in terms of the gene-environment interactions involved in modulating the preventive outcomes in epidemiologic studies and in clinical trials. The ATBC prospective randomized controlled trial of two promising nutrients exemplifies the limitations of the prior hypothesis-generating observational data and reinforces the importance of basing dietary recommendations on prospective clinical trials, whenever possible.

An important secondary finding in ATBC was the reduction in prostate cancer incidence in men receiving α-tocopherol. A secondary end point from another randomized controlled trial, which proved negative for its primary end point of testing selenium for the prevention of skin cancer, was a reduction in prostate cancers in men receiving selenium (26). Together, these secondary outcomes from two prevention trials laid the foundation for the National Cancer Institute’s second phase 3 trial for the prevention of prostate cancer. The Selenium and Vitamin E Cancer Prevention Trial (SELECT), through its factorial design, tested each nutrient separately and together in comparison to placebo. Following a recent independent review of study data, it was announced that neither supplement, taken alone or together, prevented prostate cancer and that the supplements are unlikely ever to produce the 25% disease reduction on which the study was designed. Furthermore, a small, statistically nonsignificant increase in cases of prostate cancer was observed in men older than age 50 taking only vitamin E, and a similar statistically nonsignificant increase in cases of adult onset diabetes occurred in men taking only selenium (27). This negative result of SELECT, by not confirming secondary end points in the hypothesis-generating clinical trials (26, 28), once again reinforces the importance of designing prospective clinical trials to test putative preventive agents specifically for their efficacy against given cancer end points.

In addition to confirming or refuting the benefits of specific bioactive foods during the past decade, epidemiologic and clinical studies have yielded evidence linking diet, body mass index (BMI) and obesity, and physical activity to cancer risk. In a recent systematic review and meta-analysis of 221 data sets with 282,137 incident cancer cases from prospective observational studies, BMI was linked with an increase in the risks of common and less common cancers, with associations differing between the sexes and among ethnic groups (29). Epidemiologic studies indicate the benefits of physical activity as an intervention for obesity, a known cancer risk factor, with the strongest evidence of an inverse relationship being for breast cancer, especially among postmenopausal women (30). Intervention strategies that address the effect of each lifestyle factor have been suggested by cancer prevention...
research. The second report of the World Cancer Research Fund/American Institute for Cancer Research developed lifestyle recommendations based on research in each lifestyle domain (see Table 1; ref. 31). The report used systematic literature reviews of epidemiologic and experimental studies, with assessments of the strengths and weaknesses of each study, to determine a recommendation based on the best available evidence in 20 lifestyle domains. In the same report, the World Cancer Research Fund/American Institute for Cancer Research issued recommendations on specific bioactive foods shown in epidemiologic and clinical studies to reduce cancer risk at specific sites (see Table 2). Each of the recommendations was based on the best available scientific evidence.

Although relatively new, studies of diet, BMI, and physical activity in relation to cancer prevention have yielded important findings showing promise for reducing the effect of cancer on society by encouraging healthy lifestyle choices. Increased research in these areas, especially in molecular epidemiologic approaches to basic nutritional science, should supply sound scientific data to assist policymakers in developing public health strategies.

### Reproductive Factors and Cancer Risk

The idea that hormones that control normal growth in target organs can, under the wrong circumstances, contribute to neoplastic transformation has been entertained for a long time, and is most evident in hormone-related cancers: breast, endometrium, ovary, and prostate (32). Estrogen, the best studied hormone of those implicated in carcinogenesis, is involved in several different cancer types.

The carcinogenic properties of estrogen are evident in Beatson's 1896 observation that oophorectomy is associated with the regression of breast tumors (41). The development of vaginal clear-cell adenocarcinoma following intrauterine exposure to diethylstilbestrol in daughters whose mothers were treated with this drug to prevent spontaneous abortion and premature delivery was a key early epidemiologic observation of the carcinogenic effect of estrogens (33, 34). The classic epidemiologic approach of tracking the cancers on a case-by-case basis, accompanied by detailed questioning, led to an initial report of six cases. A subsequent series of case-control studies ultimately linked this rare cancer to the synthetic estrogen (35).

Later investigations of the association of estrogen with a variety of chronic diseases offer an interesting twist to our understanding of the importance of study design for exposing true relationships between exposures and disease outcomes. Estrogen has long been viewed as a carcinogen in the breast, in part from associations noted in epidemiologic studies devoted to estimating the correlation between endogenous serum estrogen levels and breast cancer risk (36). Overall, case-control and prospective observational studies show that higher levels of endogenous serum estrogens are associated with increased breast cancer risk (36, 37). The actual estrogen exposure of the breast is presumed to be reflected in several surrogate markers of exposure (e.g., early age at menarche, older age at menopause, nulliparity, older age at first live birth, and postmenopausal obesity) for which an association with breast

### Table 1. 2007 WCRF/AICR dietary recommendations

<table>
<thead>
<tr>
<th>Recommendations for cancer prevention</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Be as lean as possible without becoming underweight</td>
</tr>
<tr>
<td>2. Be physically active for at least 30 min every day</td>
</tr>
<tr>
<td>3. Avoid sugary drinks. Limit consumption of energy-dense foods (particularly processed foods high in added sugar, or low in fiber, or high in fat)</td>
</tr>
<tr>
<td>4. Eat more of a variety of vegetables, fruits, whole grains and legumes such as beans</td>
</tr>
<tr>
<td>5. Limit consumption of red meats (such as beef, pork and lamb) and avoid processed meats</td>
</tr>
<tr>
<td>6. If consumed at all, limit alcoholic drinks to two a day for men and one a day for women</td>
</tr>
<tr>
<td>7. Limit consumption of salty foods and foods processed with salt (sodium)</td>
</tr>
<tr>
<td>8. Do not use supplements to protect against cancer</td>
</tr>
</tbody>
</table>

Special population recommendations

9. It is best for mothers to breastfeed exclusively for up to 6 months and then add other liquids and foods
10. After treatment, cancer survivors should follow the recommendations for cancer prevention


### Table 2. WCRF/AICR 2007 recommendations for cancer-preventive bioactive foods in relation to specific tissue sites

<table>
<thead>
<tr>
<th>Foods</th>
<th>Tissue site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-starchy vegetables</td>
<td>Mouth, pharynx, larynx, esophagus, stomach</td>
</tr>
<tr>
<td>Allium vegetables</td>
<td>Stomach</td>
</tr>
<tr>
<td>Garlic</td>
<td>Colorectum</td>
</tr>
<tr>
<td>Fruits</td>
<td>Mouth, pharynx, larynx, esophagus, lung, stomach</td>
</tr>
<tr>
<td>Foods containing folate</td>
<td>Pancreas</td>
</tr>
<tr>
<td>Foods containing carotenoids</td>
<td>Mouth, pharynx, larynx, lung</td>
</tr>
<tr>
<td>Foods containing β-carotene</td>
<td>Esophagus</td>
</tr>
<tr>
<td>Foods containing lycopene</td>
<td>Prostate</td>
</tr>
<tr>
<td>Foods containing vitamin C</td>
<td>Esophagus</td>
</tr>
<tr>
<td>Foods containing selenium</td>
<td>Prostate</td>
</tr>
</tbody>
</table>

Cancer risk has been noted (38). These “risk factors” have been incorporated into clinically applicable models to assess breast cancer risk level (39). Although observational studies of postmenopausal use of exogenous estrogen (hormone replacement therapy) have not consistently reported an association, current and longer duration of estrogen use is associated with increased breast cancer risk (40). This epidemiologically demonstrated association, together with early observations that oophorectomy is associated with breast tumor regression (41), stimulated the development of antiestrogenic agents—selective estrogen receptor modulators and aromatase inhibitors—for breast cancer treatment and prevention (42, 43).

In contrast to breast cancer, other estrogen-responsive chronic diseases, such as coronary heart disease (CHD), have been shown in observational studies to exhibit an inverse relationship with estrogen levels. Premenopausal women have higher levels of estrogen and a lower frequency of CHD than age-matched men (44), with the risk of CHD increasing dramatically after menopause as estrogen levels drop precipitously. Epidemiologic evidence suggested that higher endogenous estrogen levels or exogenous estrogen intake are associated with lower rates of CHD but higher risk of breast cancer. Thus, in the Nurses’ Health Study, a prospective observational cohort study of 70,533 postmenopausal women, participants with no pre-existing heart disease who were taking oral conjugated estrogen, had a decreased risk of major coronary events (45). Based on its ability to decrease levels of cholesterol, a presumed surrogate biomarker of CHD, estrogen activity has been interpreted as ameliorating CHD risk. This benefit from estrogen use observed in epidemiologic studies with regard to heart disease, together with the benefit to multiple other organ systems (e.g., bone, skin, brain, and vaginal epithelium), led prominent clinical researchers to declare that estrogen was a key antiaging drug and to encourage all menopausal women to take it, despite the absence of a prospective randomized clinical trial of this hormone for this purpose (46).

The "great estrogen conundrum" (47) pits the antithetical outcomes involving breast cancer and cardiovascular disease against each other. As the level of scientific evidence for the two disease outcomes advanced from observational to experimental, widely held views of estrogen’s overall benefits underwent major revision. The observational findings favoring estrogen use for the heart were not substantiated in prospective trials, such as the Heart and Estrogen/progestin Replacement Study (48), although the generalizability of this finding was limited by the fact that the Heart and Estrogen/progestin Replacement Study cohort consisted of women with pre-existing CHD. Publication of the Heart and Estrogen/ progestin Replacement Study in 1998 was followed by a moderate decline in hormone replacement therapy use (Fig. 3; ref. 49). The Women’s Health Initiative, a well-designed randomized controlled trial using hormone replacement therapy as a prospective intervention in healthy postmenopausal women (50), clarified the association between estrogen and these two chronic disease outcomes. The epidemiologic data regarding increased risk of
breast cancer were borne out in the Women’s Health Initiative estrogen plus progestin trial. In contrast, the observational data pertaining to CHD were contradicted (50), indicating that much of the apparent benefit seen in the epidemiologic studies was likely due to biases inherent in their weaker study design (46). Together, these two adverse disease outcomes, shown at a higher level of evidence in the Women’s Health Initiative than in the earlier, observational studies, had downstream ramifications—massive cessation of hormone replacement use by postmenopausal women, exceeding that following publication of the Heart and Estrogen/progestin Replacement Study (Fig. 3; 49). The decrease in breast cancer incidence noted 1 year later in 2003 has been attributed to this widespread change in medical practice (51), vindicating the implementation of this critically important trial, and emphasizing the necessity of implementing the highest level of study possible in addressing epidemiologic questions (46). The results of the Women’s Health Initiative emphasize the importance of adhering to the principles of evidence-based medicine in evaluating inputs, both risk-inducing and risk-reducing, in relation to disease outcomes (Fig. 4).

Prostate cancer, the most common adult cancer in men in the United States, is hormone-dependent, with its growth depending on testosterone (T) and its more active metabolite dihydrotestosterone (32). This dependency is evidenced in the successful use of androgen ablation/blockade for palliation of advanced disease. Overall, epidemiologic studies by investigators such as Ronald Ross and Brian Henderson support an association between hormone levels and prostate cancer risk, despite difficulties encountered in establishing such a link (52). In contrast, a well-conducted prospective case-control study nested in the Physicians’ Health Study showed that when levels of hormone and sex hormone-binding globulin were adjusted simultaneously, a strong trend of increasing prostate cancer risk was observed with increasing testosterone levels. Such epidemiologic evidence for testosterone involvement in prostate cancer risk, coupled with the availability of the drug finasteride, which inhibits 5α-reductase, the enzyme that catalyzes conversion of T to dihydrotestosterone, led to design of the Prostate Cancer Prevention Trial (PCPT), first large phase 3 trial of its kind (53); this trial showed a reduction in prevalent prostate cancers in men receiving the 5α-reductase inhibitor.

Occupation/Environment

Austere physician observations underlay the recognition of potential associations of occupational exposures to natural and man-made agents with cancer. Such “occupational cancers” were identified through observational epidemiologic methods as early as 1700, when Bernardini Ramazzini, an Italian physician, published De Morbis Artificum Diatriba (Diseases of Workers), a comprehensive study of the hazards of exposure to chemicals, dusts, metals, and other agents among workers in 52 occupations (54). This is
considered the first study of what would become occupational medicine. A more focused epidemiologic study was published in 1775 by Percival Pott of Saint Bartholomew’s Hospital in London, who described cancer of the scrotum in chimney sweeps, caused by constant exposure to soot (55).

Links between modern occupational exposures and cancer continue to be suggested by epidemiologic studies. In some cases, notably the links between asbestos and mesothelioma (56) and between vinyl chloride monomer and angiosarcoma of the liver (57), the rarity of the cancer facilitated the recognition of the association. Links between occupational exposures and more common cancers—such as between aromatic amines and bladder cancer (58) or between asbestos exposure and lung cancer (59)—were less obvious and were only brought to public attention through the efforts of perceptive clinicians and epidemiologists.

Quantitative assessment of exposure is difficult, posing challenges to epidemiologic investigation of occupation/cancer relationships. Whereas individuals usually can describe their smoking histories and eating habits reasonably well, they are often relatively ignorant about their work exposures. Moreover, because of long latency periods between exposure and cancer diagnosis and because working conditions change over time, data on current workplace exposures may not reflect those experienced by past workers. Epidemiologists often have had to rely on proxy measures of exposure in studies of occupational cancer. For example, the epidemiologic studies which showed that physicians who specialized in radiology prior to 1950 were at increased risk of leukemia because of high ionizing radiation exposure relied primarily on knowledge of typical radiation protection practices in different time periods, rather than on specifics about the exposure of particular individuals (60). Difficulty in obtaining accurate exposure data remains a problem today, hampering recent efforts such as the determination of whether occupational exposures to pesticides (61) and electromagnetic fields (62) influence cancer risk.

Occupational cancers represent only ~5% of all cancers in developed countries (55), and the number of cases is expected to decrease in future years because of improvements in industrial hygiene. Nevertheless, occupational cancers are significant because people will continue to suffer and die from them during the next few decades as a result of past exposures, and they are more readily preventable than cancers caused by tobacco, diet, or other lifestyle factors.

Molecular epidemiology (see below) has increased our understanding of genetic differences in susceptibility to occupational and environmental carcinogens. For example, among men at high risk of bladder cancer because of tobacco smoking and occupational exposure to aromatic amines, susceptibility to this cancer is influenced by polymorphisms in the genes for glutathione-S-transferase M1 (GSTM1), glutathione-S-transferase T1 (GSTT1), and N-acetyltransferase 2 (NAT2; ref. 63). Polymorphisms in GSTM1 and the manganese superoxide dismutase (MnSOD/SOD2) gene influence mesothelioma risk in asbestos-exposed persons (64). Differences in myeloperoxidase (MPO) genotypes modify the risk of lung cancer in asbestos-exposed workers (65). Polymorphisms in GSTT1 and in the gene for cytochrome P450 2E1 (CYP2E1) have been associated with variations in mutagenic risk in workers exposed to vinyl chloride (66). Among never-smokers, individuals who are homozygous for the GSTM1-null allele have been found to be at increased risk of developing lung cancer as a result of exposure to environmental tobacco smoke (67). Exposure to tobacco smoke through direct smoking was covered in a previous section of this review. Individual differences in factors other than genotype may also influence susceptibility to occupational and environmental carcinogens. For example, susceptibility to lung cancer associated with asbestos exposure is greatly increased by cigarette smoking (68).

One of the longest-term studies of the environmental effects on cancer is the Life Span Study of atomic bomb survivors. The Life Span Study was established to study the long-term carcinogenic effects of radiation on the survivors of the bombings of Hiroshima and Nagasaki. The cohort consists of 120,321 registered residents of Hiroshima and Nagasaki who were within 2.5 km (~54,000 survivors), between 2.5 and 10 km (~40,000 survivors), or more than 10 km (26,580 survivors) from the hypocenter at the time of bombing. Analysis of diagnoses of solid cancer between 1958 and 1998 found 17,448 first primary cancers, and ~11% were associated with atomic bomb radiation exposure (69). Radiation-associated increases in cancer rates were also found to persist throughout the survivors’ lifetimes. Forty-eight percent of the cancers among survivors receiving doses of at least 1 Gy were attributable to radiation exposure; ~18% of the excess cancers occurred among individuals in the dose range of 5 to 200 mGy. The Life Span Study provides the best quantitative risk estimates of the carcinogenic effects of radiation exposure in humans and is a major source of epidemiologic data used for radiation risk assessment and establishment of radiation protection guidelines.

### Infectious Agents

The idea that infectious agents might cause some cancers has existed for more than a century, but not until the 1950s did researchers seriously begin to evaluate such associations (70, 71). At the global level, the estimated percentage of cancers associated with infections is between 10% and 20%, with the implicated organisms ranging from viruses and bacteria to parasites, such as *Schistosoma haematobium*, a risk factor for bladder cancer (72). By the 1980s, the development of new methodologies, including PCR, facilitated documentation of associations between the presence of certain viruses and bacteria in tumor tissue and reproductive and gastric cancers, respectively (73). Using this knowledge, observational studies, such as the National Health and Nutrition Examination Survey, included molecular epidemiologic studies that helped define the breadth of infection in the U.S. population. For example, National Health and Nutrition Examination Survey data showed that almost one-half of women 14 to 44 years of age were infected with the human papillomavirus (HPV), which is responsible for practically all cases of cervical cancer (and possibly squamous cell skin and other cancers; refs. 72, 73). Population studies also suggested an association between hepatitis B and liver cancer, especially in the presence of aflatoxins in the diet (74). In addition, a population study indicated a strong link between *Helicobacter pylori* infection and family history of gastric cancer (75).

Some rarer cancers also have been associated with infectious agents. EBV, a member of the herpesvirus family, infects >95% of all humans and can cause mononucleosis in adolescents. EBV was isolated from Burkitt lymphoma, an aggressive B-cell lymphoma; although rare, Burkitt lymphoma accounts for nearly half of all childhood cancers in equatorial Africa. EBV has also been detected in nasopharyngeal carcinoma, non-Hodgkin lymphoma in AIDS patients, and T-cell and Hodgkin lymphomas. EBV is hypothesized to enhance genomic instability, leading to increased likelihood of the c-myc translocation characteristic of Burkitt lymphoma (76). In nasopharyngeal carcinoma, EBV is associated with alterations in nasopharyngeal carcinoma, EBV is associated with alterations...
in expression of the Bcl-2 and TP53 genes (77). Another member of the herpesvirus family, Kaposi sarcoma-associated herpesvirus, is consistently detected in all forms of Kaposi sarcoma and in a specific subset of lymphoproliferative disorders (78).

The human T-cell leukemia virus-1 is the etiologic agent of adult T-cell leukemia-lymphoma. Infection with human T-cell leukemia virus-1 is endemic in sites in southern Japan, the Caribbean, Central and South America, and parts of Africa and the Middle East. The virus is transmitted by breast-feeding, sexual activity, or exposure to infected blood. The viral gene tax encodes a transcriptional transactivator that immortalizes CD4+ lymphocytes, renders them resistant to apoptosis, and also induces angiogenesis (79).

Cancer prevention strategies that developed in response to epidemiologic and clinical findings on infectious agents include lifestyle changes, such as reducing exposure to infectious agents, as well as vaccination by vaccine. Vaccines for cancer prevention have received increased attention since 1981, when the hepatitis B vaccine was developed to prevent hepatitis and also prevented liver cancer (80). The etiologic association between HPV and cervical cancer had been established by the research of Harald zur Hausen dating back to the 1970s. The importance of this discovery, evidenced in zur Hausen’s receipt of the 2008 Nobel prize in Physiology or Medicine, laid the foundation for work such as that of Douglas Lowy and John Schiller, who developed an HPV vaccine consisting of the major structural viral protein L1 (virus-like particles; ref. 81). One such HPV vaccine has been tested in clinical trials, and is Food and Drug Administration–approved, and is being distributed among those at risk for cervical cancer (82, 83). Presently, a vaccine against H. pylori (84) is in early-phase testing in small numbers of people. Early results from this controlled, single-blind phase I clinical trial in 57 volunteers suggest that the vaccine is safe and potentially useful; the authors recommended further clinical study (84).

Genetic Epidemiology

A century ago, the prevailing view was that cancer was an inevitable result of an individual’s genetic destiny. By the 1960s, epidemiologic evidence suggested that up to 90% of cancers were strongly associated with avoidable factors, i.e., factors other than inherited predisposition. Recent years have seen a resurgence of research emphasis on genetic contributions. The modern field of genetic epidemiology evolved out of the recognition that some cancers cluster within families (85), implicating an inherited component in human carcinogenesis and suggesting that a similar genetic contribution might occur in nonfamilial, sporadic cancers. Parallel developments in high-throughput genotyping technologies and accompanying analytic computational methods enabled researchers to identify the actual physical locations and identities of implicated genetic factors, rather than simply establishing the existence of genetic contributions to carcinogenesis (86).

Traditional Studies Documenting Genetic Contributions to Carcinogenesis

Twin studies. An early laboratory for studying genetic contributions to complex human traits, including diseases such as cancer, was recognized in identical (monozygotic) twins. Given the shared environment in twins, differences in phenotypic concordance between monozygotic and dizygotic twins were attributed to genetic factors (87). A key study analyzing data from 11 cancer sites in 44,788 Scandinavian twin pairs revealed only a moderate increase in overall risk of the same cancer in a twin of a person with cancer, especially for breast, prostate, colorectal, lung, and stomach cancers (88). The effect of heritability (that portion of risk that is due to inherited genes) was statistically significant only for prostate (42% of risk), colorectal (35%), and breast (27%) cancer. The authors concluded that the environment was the “overwhelming contributor to the causation of cancer” in these twin populations, precipitating considerable debate (87, 89). Note, the notion that heritability fails to explain a large proportion of cancer incidence and mortality is not new, having been suggested in multiple earlier investigations (90).

The strong association of lung cancer with tobacco presents an unusual challenge to the disentanglement of genetic contributions from environmental contributions. In the Scandinavian twin study (88), lung cancer risk was increased in the twin of an individual with lung cancer, although the heritability was limited. A systematic review of case-control, cohort, and twin studies designed to assess the role of family history in lung carcinogenesis led to a similar conclusion of only a minor genetic contribution (91). The assumption that the environmental factor, smoking, is similarly concordant in monozygotic and dizygotic twins is not supported by the data, with a stronger concordance for monozygotic twins. Furthermore, men show similar lung cancer concordance ratios for monozygotic and dizygotic twins, consistent with a strong shared environmental influence.

Although twin studies offer insight into genetic contributions to cancer, the rarity of twins limits the effect of twin studies on research, necessitating reliance on alternative analytic approaches.

Family-based association studies. The aggregation of specific cancers or constellations of cancers in families is a necessary but insufficient criterion for inferring genetic contribution to disease (87). Astute clinical observations of “cancer” families led to the identification of genetic syndromes such as the Li-Fraumeni (92) and Lynch (hereditary nonpolyposis colorectal cancer) syndromes (85).

Modern Molecular Genetic Epidemiologic Studies

Genome-wide linkage analysis has successfully superimposed molecular biological tools on the pedigree approach to map the actual genes responsible for cancers with monogenic, “Mendelian,” inheritance patterns (93). Analysis of families with multiple, usually early onset cancer cases by linkage approaches such as positional cloning has identified genes that are mutated in affected individuals. BRCA1 and BRCA2 in breast and ovarian cancer syndrome, MLH1 and MSH2 in hereditary nonpolyposis colorectal cancer, and APC in familial adenomatous polyposis were identified in this manner. Such highly penetrant gene variants are implicated in only a small percentage of cancers. Most cancers are complex, quantitative traits reflecting the interplay of common variations in multiple genes (epistasis), each conferring only minor increases in risk, as well as interactions with environmental factors (93, 94).

Candidate-gene studies are hypothesis-based, involving the selection of genes and polymorphisms based on criteria such as gene function or location in a region of linkage (93). Overall, this “gene-by-gene functional candidate gene approach” (94) has had limited success; with rare exceptions, including those discussed above, few polymorphisms show consistent associations with cancer from one study to another (95). An example is a recent breast cancer candidate gene study that examined variants in 11 genes encoding proteins in the estrogen metabolic pathway (96). Although single factor analyses suggested a possible association of CYP1B1_1294_GG with risk, no significant associations were

Landmarks in the History of Cancer Epidemiology

Twin studies. An early laboratory for studying genetic contributions to complex human traits, including diseases such as cancer, was recognized in identical (monozygotic) twins. Given the shared environment in twins, differences in phenotypic concordance between monozygotic and dizygotic twins were attributed to genetic factors (87). A key study analyzing data from 11 cancer sites in 44,788 Scandinavian twin pairs revealed only a moderate increase in overall risk of the same cancer in a twin of a person with cancer, especially for breast, prostate, colorectal, lung, and stomach cancers (88). The effect of heritability (that portion of risk that is due to inherited genes) was statistically significant only for prostate (42% of risk), colorectal (35%), and breast (27%) cancer. The authors concluded that the environment was the “overwhelming contributor to the causation of cancer” in these twin populations, precipitating considerable debate (87, 89). Of note, the notion that heritability fails to explain a large proportion of cancer incidence and mortality is not new, having been suggested in multiple earlier investigations (90).

The strong association of lung cancer with tobacco presents an unusual challenge to the disentanglement of genetic contributions from environmental contributions. In the Scandinavian twin study (88), lung cancer risk was increased in the twin of an individual with lung cancer, although the heritability was limited. A systematic review of case-control, cohort, and twin studies designed to assess the role of family history in lung carcinogenesis led to a similar conclusion of only a minor genetic contribution (91). The assumption that the environmental factor, smoking, is similarly concordant in monozygotic and dizygotic twins is not supported by the data, with a stronger concordance for monozygotic twins. Furthermore, men show similar lung cancer concordance ratios for monozygotic and dizygotic twins, consistent with a strong shared environmental influence.

Although twin studies offer insight into genetic contributions to cancer, the rarity of twins limits the effect of twin studies on research, necessitating reliance on alternative analytic approaches.

Family-based association studies. The aggregation of specific cancers or constellations of cancers in families is a necessary but insufficient criterion for inferring genetic contribution to disease (87). Astute clinical observations of “cancer” families led to the identification of genetic syndromes such as the Li-Fraumeni (92) and Lynch (hereditary nonpolyposis colorectal cancer) syndromes (85).

Modern Molecular Genetic Epidemiologic Studies

Genome-wide linkage analysis has successfully superimposed molecular biological tools on the pedigree approach to map the actual genes responsible for cancers with monogenic, “Mendelian,” inheritance patterns (93). Analysis of families with multiple, usually early onset cancer cases by linkage approaches such as positional cloning has identified genes that are mutated in affected individuals. BRCA1 and BRCA2 in breast and ovarian cancer syndrome, MLH1 and MSH2 in hereditary nonpolyposis colorectal cancer, and APC in familial adenomatous polyposis were identified in this manner. Such highly penetrant gene variants are implicated in only a small percentage of cancers. Most cancers are complex, quantitative traits reflecting the interplay of common variations in multiple genes (epistasis), each conferring only minor increases in risk, as well as interactions with environmental factors (93, 94).

Candidate-gene studies are hypothesis-based, involving the selection of genes and polymorphisms based on criteria such as gene function or location in a region of linkage (93). Overall, this “gene-by-gene functional candidate gene approach” (94) has had limited success; with rare exceptions, including those discussed above, few polymorphisms show consistent associations with cancer from one study to another (95). An example is a recent breast cancer candidate gene study that examined variants in 11 genes encoding proteins in the estrogen metabolic pathway (96). Although single factor analyses suggested a possible association of CYP1B1_1294_GG with risk, no significant associations were
observed by multifactor analyses. Candidate gene approaches to lung cancer risk must necessarily genetic from environmental (tobacco) contributions (97). Thus, an XRCC1 (DNA repair gene) polymorphism shows significant association with risk overall but especially in nonsmokers, whereas heavy smokers show inverse risk (97). In addition, a recent study of gene-environment interactions identified polymorphisms in the nicotinic acetylcholine receptor gene cluster on chromosome 15q24 with effects on smoking quantity, nicotine dependence, and the risk of lung cancer and peripheral artery disease in populations of European descent (98).

The genome-wide association study (GWAS) approach is agnostic, requiring no prior knowledge of position or function of a marker (91, 99). Putative cancer-associated loci are localized by exploiting their tight linkage disequilibrium with known single nucleotide polymorphisms that are densely distributed throughout the genome and serve as surrogates for the polymorphism of interest (100). To economize on cost and the large sample size required for statistical validation of complex disease association, analysis is restricted to a subset of “tag single nucleotide polymorphisms” that capture most of the linkage in a region (100). Other cost-conserving measures include multistaged designs (sequential GWAS analyses; refs. 99–102), minor allele frequencies of ≥0.1, and odds ratios for effect sizes of ≥1.3 (94). Still, a sample size of ≥1,000 cases is required, necessitating collaboration among multiple individual projects (100).

Supporting evidence for a strong genetic component in prostate cancer, GWAS implicated the 8q24 locus in this disease, as well as colorectal, breast, and ovarian cancers (102, 103). GWAS of men aged ≤60 years at diagnosis and with a family history revealed seven loci on multiple chromosomes associated with prostate cancer, two of which (8q24 and 17q) had previously been reported (99). A GWAS study of breast cancer in high-risk Ashkenazi Jewish women who were negative for a BRCA mutation showed the strongest disease association with 6q22.33, but also a significant association with FGF2 (chromosome 10q25.3-q26; ref. 104). A Breast Cancer Association Consortium study showed strong, consistent associations of FGF2 and three other genes (TNRC9, MAP3K1, and LSP1) with breast cancer (101). In addition, lung cancer susceptibility was mapped to 15q24-25.1 by GWAS (7).

Thus, GWAS has the potential to localize cancer risk to specific chromosomal regions. Its central assumption, however, that a marker in tight linkage disequilibrium with a polymorphism that directly influences cancer risk will always be detectable in association with disease at an appropriate sample size, has been challenged (105, 106). GWAS is not self-sufficient, and embracing GWAS should not be taken as a rejection of earlier strategies. GWAS and candidate-gene investigations should proceed side-by-side. The scientific paradigm of confirmation and validation must apply to GWAS, with ultimate confirmation coming from alternate technologies. The value of GWAS will be to direct investigations to promising sites containing plausibly causative genes for candidate-gene investigations, more refined regional mapping, and detailed functional analyses.

Conclusion: Implications for Cancer Prevention

Epidemiologic research has successfully identified numerous putative and some confirmed causal factors in specific cancers. Where the strength of association is indisputable, as between tobacco and lung cancer, observational knowledge has stimulated efforts at control of the implicated factor. For this discipline to have a meaningful effect on its sister field of cancer prevention, decisions regarding the design of epidemiologic studies, including the selection of putative risk factors for evaluation, need to be made prospectively with an eye to the implications of study findings for subsequent clinical and public health research evaluating potential interventions.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

Received 9/29/2008; revised 12/8/2008; accepted 2/3/2009; published OnlineFirst 3/10/09.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The authors are grateful to Dr. Kenneth Buettow and Darrell Anderson for generous contributions to the scientific and technical preparation of this article.

References

3. Hill J. Cautions against the immoderate use of snuff. Founded on the known qualities of the tobacco plant and the effects it must produce when this way taken into the body and enforced by instances of persons who have perished miserably of diseases, occasioned, or rendered incurable by its use. Second ed. London: R. Baldwin and J. Jackson; 1761.
Landmarks in the History of Cancer Epidemiology


References for Additional Reading


Correction: AACR Centennial Series Article on Landmarks in the History of Cancer Epidemiology

In the AACR Centennial Series article on landmarks in the history of cancer epidemiology in the March 15, 2009 issue of *Cancer Research* (1), the fourth sentence of the Introduction should read "As illustrated in Figs. 1, 2, and 3, site-specific data on the mortality of women and men in the United States from cancer has been collected and published since 1930." Also, the fourth sentence of the last paragraph on page 2156 should read "Publication of the Heart and Estrogen/progestin Replacement Study in 1998 was followed by a moderate decline in hormone replacement therapy use (Fig. 4; ref. 49)." Finally, the second sentence on page 2157 should read "Together, these two adverse disease outcomes, shown at a higher level of evidence in the Women’s Health Initiative than in the earlier, observational studies, had downstream ramifications—massive cessation of hormone replacement use by postmenopausal women, exceeding that following publication of the Heart and Estrogen/progestin Replacement Study (Fig. 4; ref. 49)."

Landmarks in the History of Cancer Epidemiology

Peter Greenwald and Barbara K. Dunn


Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-09-0416

Cited articles
This article cites 99 articles, 28 of which you can access for free at:
http://cancerres.aacrjournals.org/content/69/6/2151.full#ref-list-1

Citing articles
This article has been cited by 8 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/69/6/2151.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.