The Search for Cancer Risk Factors: When Can We Stop Looking?

Colin B. Begg, PhD

A B S T R A C T

In recent decades, countless cohort, case–control, and ecologic studies have been conducted in the search for cancer risk factors. On the basis of knowledge gained from these studies, various influential commentaries have endeavored to classify the extent to which the total cancer burden is attributable to general categories of risk, such as diet, tobacco, sun exposure, and others. These commentaries have led to the conventional wisdom that most of the cancer burden is caused by environmental factors and relatively little is directly attributable to genetic susceptibility.

In the face of the apparent knowledge that the cancer burden is essentially fully "explainable" on the basis of known environmental risks, this article addresses the conceptual and empirical basis of the continued search for new risk factors. It proposes that the extent of the aggregation of cancer within individuals in the population—that is, the occurrence of second primary cancers—is a crucial statistic in this context. A study of the incidence of second primary melanoma suggests that the bulk of the risk variation in this disease cannot be explained by known risk factors. The implications of these ideas for research strategy and for public health policy are discussed. (*Am J Public Health.* 2001;91:360–364)

In 1996, the Harvard Center for Cancer Prevention published a report summarizing current knowledge regarding cancer risk, entitled the *Harvard Report on Cancer Prevention*.[1] The fundamental conclusion of this report was that "cancer is . . . a preventable illness." In support of this opinion, the report displayed a table of cancer (mortality) risks summarizing the "estimated percentage of total cancer deaths attributable to established causes of cancer." The percentages listed were tobacco (30%), diet/obesity (30%), sedentary lifestyle (5%), occupational factors (5%), family history of cancer (5%), viruses/other biological agents (5%), perinatal factors/growth (5%), reproductive factors (3%), alcohol (3%), socioeconomic status (3%), environmental pollution (2%), ionizing/ultraviolet radiation (2%), prescription drugs/medical procedures (1%), and salt/other food additives/contaminants (1%). These percentages add to 100% and thus do not accede to the possibility that there may exist other, hitherto undefined categories of risk. This presentation follows a tradition dating back to Wynder and Gori[2] in 1977 and repeated by various other commentators, notably Higginson and Muir,[3] Doll and Peto,[4] Henderson, Ross, and Pike,[5] and Ames, Gold, and Willett.[6] Notwithstanding that these numbers are presented as approximations and recognized as such by all concerned, the clear implication is that essentially all cancer risk, with the possible exception of the 5% attributed to "family history of cancer," is a result of potentially modifiable environmental risk factors. The authors of the report used these arguments to call for broad population-based cancer prevention interventions based on this knowledge. Of special importance was the recommendation to try to "shift the behavior of the whole population" rather than focus on high-risk individuals.

The essential message of the Harvard report is that we already know the major factors that cause cancer. The purpose of this article is to explore the methodological basis for this claim. This is an important issue since, if we already know the factors that cause 100% of cancer, there seems little purpose in further epidemiologic research to uncover new factors. In this article, a different conceptual structure for evaluating the totality of the influence of cancer risk factors is described, and data are presented that suggest that many of the factors influencing cancer risk have yet to be discovered.

The Attribution of Risk

The authors of the *Harvard Report on Cancer Prevention* did not provide details of their methodology for estimating the attributable percentages of risk, but it is likely that they used primarily the attributable risk. Although the term "attributable risk" does not have a consistent definition,[7,8] it is most commonly used in the cancer literature to mean the proportion by which the cancer incidence rate would be reduced if the risk factor were eliminated.

Consider a binary risk factor with a relative risk of 9 and a population prevalence of 0.5, numbers that are of an order of magnitude similar to the effects of lifetime smoking on upper aerodigestive cancers.[9] The attributable risk (A) is calculated by the formula

$$A = p(R-1)/[p(R-1)+1],$$

where p is the prevalence of the exposure and R is the relative risk. Thus, in the example, the attributable risk is 0.8. That is, we estimate that 80% of the cancer would be eliminated if the risk factor were eliminated, in

The author is with the Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY.

Requests for reprints should be sent to Colin B. Begg, PhD, Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, 1275 York Ave, Box 44, New York, NY 10021 (e-mail: beggc@mskcc.org).

This article was accepted December 1, 2000.

TABLE 1—Hypothetical Distributions of 2 Independent Cancer Risk Factors Among Cases and Controls, Each With an Attributable Risk of 0.8

Risk Factors		Population Prevalences (Controls)[a]	Incident Cases	Relative Risk
E_1	E_2			
Present	Present	0.25	0.81	81
Present	Absent	0.25	0.09	9
Absent	Present	0.25	0.09	9
Absent	Absent	0.25	0.01	1

[a]The control group in a population-based case–control study reflects the prevalences in the population of the risk factors under evaluation.

which case the entire population, by assumption, would experience the incidence rate of the 50% of the population who are not exposed. However, if 80% of the cancers are "caused" by the exposure, does this imply that only 20% remain to be explained by other risk factors? In a discussion of the issue, Rothman and Greenland refute this interpretation on the basis of logical arguments regarding causality and the fact that disease occurrence in an individual might require more than 1 causal factor.[10] However, the fallacy can be demonstrated very simply by counterexample on a purely empiric basis. Suppose that there exists another risk factor denoted E_2 that is uncorrelated with the first risk factor (E_1) but that also has a prevalence of 0.5 and a relative risk of 9. The resulting distribution of these factors among cases and population controls is then as displayed in Table 1. In this hypothetical scenario, each risk factor has an attributable risk of 80%, and one might conclude (erroneously) that together they "cause" 160% of the cancer risk.

So where is the flaw in this argument? The calculation of attributable risk requires a "baseline" category of risk, representing those unexposed to the risk factor. In calculating the attributable risk for E_1 we use as the baseline the subjects who are unexposed to E_1, that is, those classified in the bottom 2 rows of the table. However, these 2 groups differ in risk by a factor of 9 and so do not enjoy a homogeneous baseline risk. Clearly, the lowest-risk subjects are in the group unexposed to either factor, in the fourth row of the table. One can calculate correctly the proportion of cases attributable to both factors simultaneously by evaluating the reduction in incidence that would occur if all of the population experienced the same risk as this jointly nonexposed group. In this case, the appropriate formula is

$$A = 1 - (p_1R_1 + p_2R_2 + p_3R_3 + p_4R_4)^{-1},$$

where the p's represent prevalences (in the controls) and the R's represent relative risks with respect to the baseline.[11] That is,

$$A = 1 - [(0.25 \times 81) + (0.25 \times 9) + (0.25 \times 9) + (0.25 \times 1)]^{-1} = 0.96,$$

and so the 2 factors jointly explain 96% of the risk.

Clearly, the attributable risk of 1 risk factor does not provide evidence for or against the presence of other risk factors. Thus, the attributable risk is not immediately useful in addressing the question in the title of this article. Despite this, the presence of high attributable risks has frequently been used in the past as evidence against the possibility that direct genetic predisposition is responsible for more than a very small percentage of the cancer burden. The notion that genetic susceptibility, in isolation, causes around 2% to 5% of human cancer can be traced to Knudson,[12] and the figure of 5% has been quoted often (see, for example, Perera,[13] Harris,[14] and many others). To provide motivation for the recent explosion of interest in investigating and searching for new genetic risk factors for cancer, the justification has been advanced that the hypothetical genetic effects act in concert with known environmental factors, via "gene–environment interactions."[15,16] However, the example in Table 1 shows that the search for genetic susceptibility need not be restricted to interactive effects, since strong independent genetic factors can still be present even when the attributable risk of known environmental effects is very high.

The Concept of Risk

Cancer risk may be influenced by factors external to the subject (i.e., environmental risk factors), by genetic susceptibility, or by the interplay of both of these (i.e., gene–environment interactions). Each individual in the population has a unique risk, and a primary purpose of epidemiologic research is to determine the extent to which this risk varies from person to person and the factors that explain this variation. Each person's immediate risk also changes markedly as the person ages. Prediction of which individuals will be diagnosed with cancer in the future may also be refined by clinical or laboratory features that reflect an existing undetected cancer or biological changes on the causal pathway to a future cancer—for example, the presence of premalignant lesions such as oral leukoplakia or elevated prostate-specific antigen (PSA) levels. In all of the discussion that follows, these latter factors are not considered to be "risk" factors. They are markers of relevant biological events that have already occurred, possibly influenced by risk factors.

Individualistic risk prediction has become a focus of attention since publication of the "Gail model" for predicting breast cancer risk,[17] in which individual risks are predicted on the basis of 4 breast cancer risk factors in addition to age. The essential question addressed in this article is, Are these individual risk predictions accurate on a person-by-person basis, or can they be refined by knowledge of additional risk factors? In fact, we now know that there exists strong genetic susceptibility for individuals with mutations in the *BRCA1* or *BRCA2* genes, factors not known when the Gail model was developed. Thus, for example, if a woman is predicted by the Gail model as having a 10% risk of breast cancer in the next 30 years, this risk is truly much higher than 10% if the woman also has a *BRCA1* mutation and somewhat lower than 10% if she does not. In other words, the actual risks of women with identical Gail "scores" will vary around this average risk, and this variation will be due to additional, possibly unknown risk factors, including *BRCA1* and *BRCA2* mutational status.

The Population Distribution of Risk

How can we determine whether a group of individuals with apparently similar risks actually possess the same risk, or whether disease occurrence can be predicted more accurately on the basis of additional, hitherto unknown risk factors? This question can be considered as having a solution between 2 extremes: at one extreme, disease occurrence can be predicted with perfect accuracy; at the other extreme, the occurrence is completely stochastic (i.e., random), and no single individual in the risk group actually has a greater or lesser chance of experiencing the disease than any other at the outset.

Replication offers the theoretical solution to this problem. We cannot replicate the life experience of an individual with a unique risk, but we can conduct studies that shed light on this issue. For example, studies of the occurrence of disease in identical twins allow us to explore the extent to which genetic risk varies between sets of twins. The degree to which disease aggregates within twins is an approximate

representation of the degree to which the inherent genetic risk varies from twin pair to twin pair. For example, if there were no genetic component to risk, the risks of disease in the twins of diseased probands should be unremarkable. That is, they should be representative of the risks in all twin pairs, and thus the observed incidence rate among twins of diseased probands should be similar to the rate of occurrence of disease across all twin pairs. In fact, it can be shown that the ratio of these 2 rates is directly related to the coefficient of variance of the individual genetic risks.[18]

Unfortunately, studies of twins have limited usefulness in cancer owing to the infrequency of identical twinning and the rarity of occurrence of most cancers, although a number of twin studies of this nature have been conducted,[19,20] including a comprehensive recent study using Swedish, Danish, and Finnish twin registries in which substantial aggregations of all the major cancers in twin pairs were observed.[21] A related strategy that has the potential to provide access to much larger numbers of subjects is to study the occurrence of multiple primary cancers in the same individual. If we accept the premise that a second primary cancer is a truly independent occurrence of the disease (as opposed to a metastasis or a tumor derived from a "clone" that led to the first primary), then the evaluation of the patterns of incidence of second primaries is, in effect, a paired study, akin to a twin study, in which each subject is used as his or her own control. The standardized incidence ratio (SIR), suitably stratified by age, is the ratio of the incidence rate of (second) primaries in individuals in which a first primary has already occurred to the underlying age-specific incidence rate in the population. The magnitude of this ratio is a direct measure of the variation in the person-to-person risk in the population, with the recognition that this variation encompasses all environmental risks in addition to genetic risks, since a single individual is matched with himself or herself with respect to both genetic risks and environmental exposures.

The study of the incidence of second primary cancers thus provides us with an avenue for evaluating the degree of risk variation in the population. This strategy is at best very approximate, as there are numerous influences that could disrupt the fundamental assumption that in a single individual the probability of occurrence of a second primary in a defined time interval is the same as the probability (risk) that gave rise to the first primary. These include the fact that patients with cancer receive continuing diagnostic scrutiny at a greater intensity than the population at large, the possibility that the treatment for the first primary may affect the risk of the second, the possibility that risk factors also affect prognosis, the possibility that risk may be episodic, and others.[22] Also, the fact that any individual's second primary must occur at a more advanced age than the individual's first primary necessitates careful age stratification. Nonetheless, the age-stratified SIR can provide a first approximation of the coefficient of variation of risk in the population. In a similar vein, the SIR representing co-occurrence of 2 cancers of different types reflects the extent to which the risks of the 2 cancers are correlated in the population.[23]

Our ability to evaluate the risk variation in this way for a single type of cancer is most seriously hampered in practice by the fact that surgery for the first primary will often eliminate much of the organ tissue that would be "at risk" for a second primary, rendering the incidence rates of second primaries uninterpretable. However, melanoma is a particularly suitable model for study, since only a very small proportion of a subject's skin is removed at surgery, leaving essentially all of the skin "at risk" for a second primary. For breast cancer, contralateral occurrences of the disease are considered second primaries by convention, and it is reasonable to suppose that half of the organ tissue—1 breast rather than 2 breasts—remains at risk for a second primary. The appropriate calculation of the SIR thus involves doubling the observed rate of incidence of contralateral breast cancer. In Table 2, the SIRs are presented in 10-year age intervals for these 2 cancers, with data from the Surveillance Epidemiology and End Results (SEER) Program of the National Cancer Institute (NCI). The results are restricted to White females. The aggregated results in the bottom row indicate a high overall SIR of 8.5 for melanoma and a somewhat lower value of 3.9 for breast cancer. This latter value is not dissimilar to estimates of 3.1 and 5.2 obtained from studies of monozygotic twins.[19,21] However, the table shows that these aggregates disguise a strong age trend, especially for breast cancer. As is shown in the next section, these numbers represent large variations in risk in the population.

Contributions of Known Risk Factors

The contribution of an individual known risk factor to the population variation in risk, or the combined contributions of several risk factors, is estimable with data obtained from a case–control study conducted in the same population.[22] In fact, the contribution of a set of risk categories to the SIR is given by the formula

$$(p_1R_1^2+p_2R_2^2+\cdots)/(p_1R_1+p_2R_2+\cdots)^2,$$

where the p's and R's represent the prevalences and relative risks of a set of mutually exclusive categories, as in the example in Table 1.

TABLE 2—Standardized Incidence Ratios (SIR) of Second Primary Cancers: Melanoma and Breast Cancer

Age,[a] y	Second Melanoma, No. (95% CI)	Contralateral Breast,[b] No. (95% CI)	Breast–Melanoma,[c] No. (95% CI)	Melanoma–Breast,[d] No. (95% CI)
<30	23.0 (12) (12, 40)	218.0 (12) (113, 381)	. . . (0) (. . .)	. . . (1) (. . .)
30–39	8.9 (26) (5.8, 13.0)	31.0 (225) (27, 36)	1.4 (6) (0.5, 3.0)	1.3 (12) (0.7, 2.3)
40–49	9.2 (41) (6.6, 12.5)	7.4 (643) (6.8, 8.0)	1.3 (27) (0.9, 1.9)	1.3 (47) (1.1, 1.7)
50–59	5.3 (23) (3.4, 8.0)	4.4 (1054) (4.1, 4.6)	1.5 (62) (1.1, 1.9)	1.5 (76) (1.2, 1.9)
60–69	6.6 (29) (4.4, 9.5)	3.8 (1723) (3.7, 4.0)	1.2 (68) (0.9, 1.5)	1.2 (87) (1.0, 1.5)
70–79	11.5 (37) (8.1, 15.8)	3.4 (1596) (3.2, 3.6)	1.3 (73) (1.0, 1.7)	1.5 (88) (1.2, 1.9)
≥80	9.1 (19) (5.5, 14.3)	2.9 (861) (2.7, 3.1)	1.4 (56) (1.1, 1.9)	1.1 (35) (0.8, 1.5)
Overall	8.5 (187) (7.4, 9.9)	3.9 (6114) (3.8, 4.0)	1.3 (292) (1.2, 1.5)	1.3 (345) (1.2, 1.5)

Note. All statistics are based on rates for White females. CI = confidence interval.
Source. Data obtained from the Surveillance Epidemiology and End Results (SEER) Program of the National Cancer Institute, 1973–1994.
[a]Age at diagnosis of second primary.
[b]Results are restricted to second primaries in the contralateral breast, and the SIR is doubled to account for the fact that the number of cells at risk in the contralateral breast is approximately half of the number at risk for the first primary.
[c]Melanoma following breast cancer.
[d]Breast cancer following melanoma.

This contribution is 1 (i.e., no contribution) only if all of the relative risks are 1. In Table 1, the contribution of E_1, based on a prevalence of 0.5 and a relative risk of 9, is 1.64, that is,

$$[(0.5\times9^2)+(0.5\times1^2)]/[(0.5\times9)+(0.5\times1)]^2.$$

This is also the contribution of E_2. The combined effect of these 2 independent risk factors is 2.69. In general, the joint contribution of 2 factors is attenuated if they are not statistically independent or if their combined effects are not multiplicative.

A formal analysis of this nature has been performed to evaluate the known risk factors for melanoma in White subjects.[22] Data from a case–control study of melanoma conducted in Connecticut were used to evaluate the contributions of established risk factors: sun exposure, number of nevi, light skin color, inability to tan, light eye color, light hair color, and tendency to freckle. A logistic regression model was used to characterize the joint effects of these factors.[24] When the resulting contribution of these known risk factors was compared with the observed SIRs of second primary melanoma, obtained with data from the SEER registry, the risk factors appeared to "explain" only 23% of the population variation in risk for women in the age group with the lowest SIR, and they explained only 5% of the variance for the youngest age group. Recent research has identified a highly penetrant susceptibility gene (p16) for melanoma,[25–27] and undoubtedly this factor must be responsible for at least some of the unexplained variation in risk. However, the prevalence of p16 is believed to be very low. For example, for a gene with a population prevalence of 0.1% (1 in 1000), even if the relative risk is as high as 20, the contribution to the SIR is only 1.35. It is in this context that observed SIRs ranging from 5.3 to 23 must be interpreted.

Discussion

We have seen that the common practice of using attributable risks in evaluating the combined impact of cancer risk factors, with the attributable percentages adding to 100% or thereabouts, is seriously flawed. This practice is frequently used to conclude that the percentage of cancer explained by genetic susceptibility is low.[1–6,15,28] We have seen via a simple counterexample that this conclusion is erroneous. It is entirely possible that the cancer burden is caused primarily by genetic susceptibility, notwithstanding our knowledge of environmental and lifestyle risk factors and their attributable risks.[18,29]

In part, the widespread belief that the role of environmental factors greatly exceeds the role of genetic factors is based on the huge international variations in cancer rates[30] and the fact that risks in migrant populations gradually come to resemble those of the adopted environment.[31] The measure of population variation in risk advocated in this article is a population-specific statistic and thus is solely dependent on the variations in risk exhibited by individuals within the index population. If we were to define the population on an international basis, then the total population variation in risk would increase to the extent that there would exist systematic variations in risk between populations. The fact that variations in environmental exposures within a country may be relatively small has been used as an explanation of why, for example, case–control studies of dietary factors conducted in the United States demonstrate relatively small and inconsistent effects.[32]

Research strategy should be influenced by consideration of these issues. We need to have a conceptual strategy for determining when all the relevant risk factors have been established and when the search for new risk factors is likely to be fruitless. Recently, a paradigm shift has been witnessed in cancer epidemiology, spurred by the discovery of major susceptibility genes and by the rapidly developing technology for conducting genetic studies. A major thrust of this research has been the search for gene–environment interactions.[16] The focus on gene–environment interactions appears to be motivated, in part, by the belief that the discovery of major susceptibility genes that act independently of environmental hazards is unlikely, owing to the prevailing wisdom that all but a small proportion of cancer incidence is "explained" by known environmental agents. However, we have seen that there is no logical or empirical basis for this belief. That is, cancer risk may be influenced by numerous low-penetrance genetic abnormalities that confer their contributions to risk independently of each other and independently of environmental risk factors.

A principal thesis of this article is that careful study of the incidence rates of multiple primary cancers can be informative in motivating research strategy in the search for new risk factors, and indeed the study of multiple primaries is a burgeoning area of research.[33] Sites where multiple primaries are relatively common are suggestive of a concentration of risk in the individuals who experience the multiple primaries. In anatomic regions such as the upper aerodigestive tract, the increased occurrence of multiple primaries in different sites is clearly due in part to a common environmental risk factor, cigarette smoking.[34] This phenomenon has led to theories of "field cancerization," whereby the presence of multiple tumors is viewed as a manifestation of a single occurrence of the disease.[35] However, strong predisposition in the individual is an alternative explanation, and careful tissue evaluation by new molecular pathologic tools may ultimately assist in distinguishing distant clones from independent occurrences of the disease.[36] The study of the co-occurrence of pairs of cancers on a population basis can also provide clues to the presence of common etiology. In fact, if we examine the last 2 columns of Table 2, we observe that the risks of melanoma following breast cancer, and of breast cancer following melanoma, are both modestly elevated and unrelated to age, suggesting the absence of highly penetrant risk factors that are common to both cancers.

There are several important limitations to the interpretation of the SIRs of second primaries as precise measures of population risk variation, as outlined earlier. A particular concern is the assumption that factors that affect risk do not also affect prognosis. For example, if a risk factor also leads to shortened survival in cancer patients, this will cause the SIR to underestimate the true variation in risk in the population, and vice versa. This problem, and the other potential biases described earlier, dictate that we must interpret the observed SIRs with caution and view them as grossly representative of risk variation.

Finally, the ideas presented have implications for public health policy. The predominant approach to translating results from cancer epidemiologic research into policy for reducing cancer incidence and mortality has been broadly population based. Influential organizations such as the American Cancer Society and the NCI have encouraged broad changes in behavior, such as reduced sun exposure, diets with lower fat and higher consumption of fruits and vegetables, widespread mammography screening, and so forth. In fact, the *Harvard Report on Cancer Prevention* explicitly advocates "broad-scale interventions" in preference to a focus on "individuals defined as being at high risk."[37] Recognizing that success in influencing public behavior is a complex issue, the motivation to engage in broad campaigns vs those based on a focused "high-risk" strategy should be influenced, at least in part, by the breadth of the risk distribution in the population. If it can be shown that the great majority of the population enjoys very low risk, with the bulk of the risk concentrated in a small subset, then the merit of a focused prevention strategy is enhanced. Our study of melanoma suggests a very broad risk distribution across all age groups. Our knowledge base in identifying the high-risk subjects is limited at present, but as the determinants of risk become better understood, and our risk projections become more accurate, the rationale for a focused prevention strategy can only increase. □

Acknowledgments

I am grateful to Brian Henderson, Charlie Poole, and Alfredo Morabia for constructive feedback on earlier drafts of this manuscript, to Laura Cramer for assistance with the calculations of incidence rates, and to Marianne Berwick and Jaya Satagopan for influencing the development of the ideas presented.

References

1. Colditz G, DeJong W, Hunter D, Trichopoulos D, Willett W, eds. Harvard Report on Cancer Prevention. Vol 1. *Cancer Causes Control.* 1996; 7(suppl):S3–S59. Vol 2. *Cancer Causes Control.* 1997;8(suppl):S1–S50.
2. Wynder EL, Gori GB. Contribution of the environment to cancer incidence: an epidemiologic exercise. *J Natl Cancer Inst.* 1977;58: 825–832.
3. Higginson J, Muir CS. Environmental carcinogenesis: misconceptions and limitations to cancer control. *J Natl Cancer Inst.* 1979;63: 1291–1298.
4. Doll R, Peto R. *The Causes of Cancer.* New York, NY: Oxford University Press; 1981.
5. Henderson BE, Ross RK, Pike MC. Toward the primary prevention of cancer. *Science.* 1991; 254:1131–1138.
6. Ames BN, Gold LS, Willett WC. The causes and prevention of cancer. *Proc Natl Acad Sci U S A.* 1995;92:5258–5265.
7. McMahon B, Pugh TF. *Epidemiology: Principles and Methods.* Boston, Mass: Little, Brown; 1970.
8. Schlesselman JJ. *Case–Control Studies.* New York, NY: Oxford University Press; 1982.
9. American Cancer Society. *Cancer Facts and Figures—1998.* Atlanta, Ga: American Cancer Society; 1998:1.
10. Rothman KJ, Greenland S. Causation and causal inference. In: Rothman K, Greenland S, eds. *Modern Epidemiology.* 2nd ed. Philadelphia, Pa: Lippincott-Raven; 1998:7–28.
11. Walter SD. The estimation and interpretation of attributable risk in health research. *Biometrics.* 1976;32:829–849.
12. Knudson AG. Genetic predisposition to cancer. In: Hiatt HH, Watson JD, Winsten JA, eds. *Cold Spring Harbor Conferences on Origins of Human Cancer.* Vol 4. Cold Spring Harbor, NY: Cold Spring Harbor Laboratories; 1977:45–52.
13. Perera FP. Molecular epidemiology: insights into cancer susceptibility, risk assessment, and prevention. *J Natl Cancer Inst.* 1996;88:496–509.
14. Harris EL. Importance of heritable and nonheritable variation in cancer susceptibility: evidence from a twin study. *J Natl Cancer Inst.* 1997;89: 270–272.
15. Strong LC, Amos CI. Inherited susceptibility. In: Schottenfeld D, Fraumeni J, eds. *Cancer Epidemiology and Prevention.* 2nd ed. New York, NY: Oxford University Press; 1996: 559–583.
16. Perera FP. Environment and cancer: who are susceptible? *Science.* 1997;278:1068–1073.
17. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for White females who are being examined annually. *J Natl Cancer Inst*. 1989;81: 1879–1886.
18. Easton D, Peto J. The contribution of inherited predisposition to cancer incidence. *Cancer Surv.* 1990;9:395–416.
19. Thomas DC, Langholz B, Mack W, Floderis B. Bivariate survival models for analysis of genetic and environmental effects in twins. *Genet Epidemiol.* 1990;7:121–135.
20. Ahlbohm A, Lichtenstein P, Malmstrom H, Feychting M, Hemminki K, Pedersen NL. Cancer in twins: genetic and nongenetic familial risk factors. *J Natl Cancer Inst.* 1997;89: 287–293.
21. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer: analyses of cohorts of twins from Sweden, Denmark and Finland. *N Engl J Med.* 2000;343:78–85.
22. Begg CB, Satagopan JM, Berwick M. A new strategy for evaluating the impact of epidemiologic risk factors for cancer with application to melanoma. *J Am Stat Assoc.* 1998;93: 415–426.
23. Begg CB. Methodological and statistical considerations in the study of multiple primary cancers. In: Neugut AI, Meadows AT, Robinson E, eds. *Multiple Primary Cancers.* Philadelphia, Pa: Lippincott Williams & Wilkins; 1999: 13–25.
24. Berwick M, Begg CB, Fine JA, Roush GC, Barnhill RL. Screening for cutaneous melanoma by skin self-examination. *J Natl Cancer Inst.* 1996;88:17–23.
25. Hussussian CJ, Struewing JP, Goldstein AM, et al. Germline p16 mutations in familial melanoma. *Nat Genet.* 1994;8:15–21.
26. Kamb A, Shattuck-Eidens D, Eeles R, et al. Analysis of the p16 gene (CDKN2) as a candidate for chromosome 9p melanoma susceptibility locus. *Nat Genet.* 1994;8:23–26.
27. Wainwright B. Familial melanoma and p16—a hung jury. *Nat Genet.* 1994;8:3–5.
28. Vogelstein B, Kinzler KW, eds. Preface. In: *The Genetic Basis of Human Cancer.* New York, NY: McGraw-Hill; 1998:xv–xix.
29. Begg CB. Genetic and environmental causes of cancer [letter]. *N Engl J Med.* 2000;343:1495.
30. Ries LAG. International range of cancer incidence. In: Harras A, ed. *Cancer Rates and Risks.* Bethesda, Md: National Cancer Institute, National Institutes of Health; 1996:24–28. NIH publication 96-691.
31. Thomas DB, Karagas MR. Migrant studies. In: Schottenfeld D, Fraumeni J, eds. *Cancer Epidemiology and Prevention.* 2nd ed. New York, NY: Oxford University Press; 1996: 236–254.
32. Wynder EL. Invited commentary: response to *Science* article, "Epidemiology faces its limits." *Am J Epidemiol.* 1996;143:747–749.
33. Neugut AI, Meadows AT, Robinson E, eds. *Multiple Primary Cancers.* Philadelphia, Pa: Lippincott Williams & Wilkins; 1999.
34. Begg CB, Zhang ZF, Sun M, Herr HW, Schantz SP. Methodology for evaluating the incidence of second primary cancers with application to smoking-related cancers from the Surveillance Epidemiology and End Results (SEER) program. *Am J Epidemiol.* 1995;142: 653–665.
35. Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium. *Cancer.* 1953;6:963–968.
36. Califano JA, Sidransky D. Molecular analysis of tumors. In: Neugut AI, Meadows AT, Robinson E, eds. *Multiple Primary Cancers.* Philadelphia, Pa: Lippincott Williams & Wilkins; 1999: 225–232.
37. Colditz G, DeJong W, Emmons K, Hunter DJ, Mueller N, Sorenson G. Summary. *Cancer Causes Control.* 1997;8(suppl):S47–S50.