Published in final edited form as:
Int J Cancer. 2011 August 15; 129(4): 931–937. doi:10.1002/ijc.25714.

A Strategy for Distinguishing Optimal Cancer Sub-Types

Colin B. Begg
Memorial Sloan-Kettering Cancer Center, Department of Epidemiology and Biostatistics, 307 East 63rd Street, New York, NY 10065, Phone: (646) 735-8108, Fax: (646) 735-0009, beggc@mskcc.org

Abstract

Much attention is directed currently to identifying sub-types of cancers that are genetically and clinically distinct. The expectation is that sub-typing on the basis of somatic genomic characteristics will supplant traditional pathological sub-types with respect to relevance for targeted therapies and clinical course. Less attention has been paid to the goal of validating sub-types on the basis of the distinctiveness of their etiologies. In this article it is shown that studies of individuals with double primary malignancies provide uniquely valuable information for establishing the etiologic distinctiveness of candidate tumor sub-types. Studies of double primaries have the potential to definitively rank candidate taxonomic systems with respect to their etiological relevance by determining which sub-types are most highly correlated in the double primaries. The concept is illustrated with data from studies of the concordance of estrogen and progestin status in bilateral breast cancers, where it is shown that double primaries are much more likely to be concordant with respect to ER status than for PR status. The high concordance of ER status is consistent with a growing literature demonstrating the etiologic distinctiveness of ER+ and ER- tumors.

Keywords

Tumor sub-types; double primary cancers; etiologic heterogeneity

INTRODUCTION

Much attention is currently being directed at the goal of classifying cancers into distinct molecular sub-types, using extensive genomic data that distinguishes the somatic characteristics of these sub-types.[1,2] It is anticipated that sub-classifications based on molecular characteristics will lead to a better understanding of cancer biology, and better avenues for determining appropriate targeted therapies.[3,4] Efforts to validate the relevance of candidate sub-types have usually focused on establishing clinical distinctiveness on the basis of, say, different survival for patients in the sub-types, or by merely showing that the genomically determined sub-types differ systematically with respect to conventional pathologic criteria. It is reasonable to speculate, however, that sub-types that are genuinely biologically distinct are likely also to possess distinct etiologies. Conventionally, epidemiologists have examined the potential etiologic heterogeneity of cancer sub-types by comparing the candidate sub-types with respect to the presence of known risk factors for the cancer in question.[5] Since it is widely believed that many risk factors for cancer, especially genetic factors, have not yet been identified, this strategy is necessarily limited in scope. The

Novelty: The article provides a unique technical strategy to determine if candidate tumor sub-types are etiologically distinct.
Impact: The method has the ability to determine, among candidate tumor classification systems, which system is the most relevant for defining sub-types that are etiologically distinct.

thesis of this article is that optimal sub-typing of cancer from an etiologic perspective can be accomplished without the need to know any risk factors, merely by studying the co-occurrence of the sub-types in series of patients with double primary malignancies.

The ageing of the population, allied to a gradual increase in survival following cancer has led to a rapid increase in the occurrence of second primary cancers. In fact, over 16% of all cancers reported to the SEER registries in the USA are second primaries.[6] This has led to recognition of second primaries as a special resource for epidemiologic studies, especially where the two primaries have occurred in the same organ.[7–10] A unique feature of the occurrence of a double primary is the opportunity it affords for comparison of characteristics of the two tumors. Thus investigators have been interested in whether independently occurring cancers in an individual patient are more likely to be similar with regard to tumor pathologic characteristics. For example, many investigators have endeavored to correlate the estrogen receptor (ER) status of pairs of contralateral breast cancers.[11–21] These studies have generally shown a fairly high correlation. That is, if the patient's first cancer is ER+ then the second cancer is much more likely to be ER+ as well. This has led to speculation about the etiologic causes of this aggregation. But this is hampered by absence of a conceptual structure for interpreting these results in the context of cancer risk.

In this article a conceptually simple mathematical structure for interpreting the results of studies of tumor-sub-classifications between double primaries is presented, allowing unique insight into the etiologic heterogeneity of tumor sub-classifications. A framework is provided for comparing the relative merits of competing classification systems, such as those based on traditional gross pathology and systems based on, for example, selected tumor markers or clustering of genome-wide array patterns. The results also provide insight into study design strategies for identifying new cancer risk factors, and for interpreting the results of existing case-control studies, and as a basic framework for interpreting the interplay of germ-line and somatic genomic profiles.

MATERIALS AND METHODS

The fundamental premise is that the degree of correlation of the sub-types of independent double primary tumors is directly related to the degree of risk heterogeneity in the population with respect to the sub-types. This relationship can be expressed by a relatively simple equation, but first we need to define and explain the terms that are used.

Measuring correlation in tumor sub-types

Consider for simplicity two tumor sub-types, A and B. The relevant data available from a study of double primaries is a simple cross-tabulation of the frequencies of occurrence of these sub-types in the two tumors from each patient. Let these frequencies be denoted by F_{AA}, F_{BB}, F_{AB} and F_{BA} (Table 1), where for example F_{AA} is the number of patients in which both tumors are of type A, F_{AB} is the number of patients where the first tumor is of type A and the second tumor is of type B, etc.. For reasons that are explained in detail in the Statistical Appendix the key measure of association is the odds ratio. The odds ratio from the 2×2 table cross-classifying the tumor types is defined as the parameter ψ, and its estimate is given by

$$\widehat{\psi}=\frac{F_{AA}F_{BB}}{F_{AB}F_{BA}}. \qquad (1)$$

Measuring risk heterogeneity in the population

Our goal is to relate the preceding odds ratio to the risk heterogeneity of the two sub-types. Conceptually, each individual in the population can be classified into one of a large number of risk categories, where people in a risk category all possess similar risk of the disease under investigation. On this basis we can define terms that characterize the variation in risk among individuals in the population. Specifically, we can define K^2 to be the population coefficient of risk variation for the cancer type overall. This is the variance of the risks of individuals in the population divided by the square of the mean risk. We can define analogous terms that characterize the variation in the risks for each of the tumor sub-types, and the degree by which these risk profiles are aligned. Specifically, we define K_A^2 to be the population coefficient of risk variation for the cancer type A, we define K_B^2 to be the coefficient of risk variation for tumor sub-type B, and we define K_C to be a corresponding standardized term for the risk covariance, representing the degree to which risks of A and B are aligned from person to person. Thus K_C is a natural term for characterizing the degree of risk heterogeneity between tumor sub-types A and B. These terms represent the true variations in risk among individuals in the population, though they cannot be observed directly without complete and perfect knowledge of all of the factors that influence risk, both known and unknown.

Relationship between concordance of tumor sub-types and etiologic heterogeneity

The key result is that the observed association (odds ratio) between tumor types among double primaries is directly related to the degree of risk covariation in the population. Specifically,

$$\psi=\frac{(1+K_A^2)(1+K_B^2)}{(1+K_C)^2} \tag{2}$$

That is ψ is inversely related to K_C, and so ψ can equivalently be used as a measure of risk heterogeneity with the advantage that, unlike K_C it can be observed directly. The conceptual simplicity of this result relies on some key approximating assumptions. The two tumors in each patient must be biologically independent, an assumption that is degraded to the extent that metastases are mis-diagnosed as second primaries. It is also assumed that the two cancer occurrences are experimental replicates, driven by an identical constellation of genetic and environmental risk factors possessed by the patient, an assumption that is necessarily an approximation. Further discussion of the credibility of these and other assumptions, along with a proof of the mathematical result, is provided in the Statistical Appendix.

Conceptual interpretations

To understand the implications of equation (2), consider some special cases. First consider the circumstance in which the risk profiles of the two sub-types are perfectly correlated. This is the case if, for every risk category, the risk of tumor type A is directly proportional to the risk of tumor type B. In this setting it is easily shown that $\psi = 1$. In other words, independence in the occurrence of A and B tumors (an odds ratio of 1) corresponds to perfect alignment (a perfect correlation of 1) of their risk profiles. The extreme opposite would occur in the improbable context of risk exclusivity. This occurs if a person who has a positive risk for tumor A necessarily has zero risk of tumor B, and vice versa. In this case the odds ratio is infinite. This result can be understood by recognizing that risk exclusivity would imply that only double primaries of type AA or type BB could occur, since no person is at risk for both A and B tumors, and so the denominator of the odds ratio in equation (1) is inevitably zero. The more important point to recognize is that the more negatively correlated the population risk profiles, i.e. the more heterogeneous the risk profiles, the greater the odds ratio observed in the tumor types of double primaries. Finally, consider the situation in which the risks of the two tumor sub-types are simply uncorrelated in the population, i.e. the linear correlation coefficient is zero. In this case $\psi=(1+K_A^2)(1+K_B^2)$. In this setting the magnitude of the odds ratio between tumor sub-types is determined by a combination of the underlying degrees of risk variation of the two tumor sub-classifications.

Example: Breast cancer sub-types

These concepts can be illustrated in the context of breast cancer where numerous studies correlating pathologic characteristics of bilateral breast cancers have been published. A particular aspect of breast tumors that has garnered a lot of attention is the concordance of hormone receptor levels. A literature search was undertaken to identify articles in which cross-tabulated frequencies of either ER or PR status, or both, were reported in women with contra-lateral breast cancer. This involved initially a Medline search using the key words “receptor” “contralateral” and “breast”. This was followed up by examining articles cited in those identified in the Medline search. This process identified 10 articles in which data were presented in sufficient detail for our purposes.[11–20] An especially large study was identified where the cross-tabulated data were not presented, and these data were kindly supplied by the author upon request.[21] The results were initially evaluated for heterogeneity using the Monte Carlo version of the Breslow and Day statistic, and based on these results summary odds ratios and confidence intervals were calculated using the Mantel-Haenszel method.

RESULTS

The data are presented in Table 2. A strong association for ER receptor status is identified, with a summary odds ratio of 5.2 [3.8–7.2]. In contrast, the association of PR status in the two tumors is seen to be much more modest (OR=2.1, 95% CI 1.6–2.9). These results indicate that the ER status of breast cancers has etiologic relevance, and that sub-groups of breast cancers characterized by ER status should possess distinct risk factors. Conversely, there is considerably less evidence to support the hypothesis that classifying breast cancers by PR status would be likely to be fruitful in identifying distinct risk factors. Note that the study by Weitzel et al.[20] was excluded due to the fact that the observed odds ratio was substantially higher than the other studies, and because this study was conducted solely in *BRCA1/2* carriers. Inclusion of this study would further accentuate the distinction observed between the summary odds ratios for ER and PR.

DISCUSSION

The explosion of information about the extent and nature of somatic mutations in cancers during the past several years has led to a renewed interest in the pathologic sub-classification of tumors. Many studies have been conducted that endeavor to use, for example, genome-wide data to sub-classify tumors on the basis of their somatic molecular characteristics.[22,23] These molecular classifications have challenged the notion that traditional pathologic criteria are the most relevant for classifying tumors. Tumors can be classified on the basis of the presence of a somatic mutation in a single gene, such as TP53, or a small number of distinct candidate genes. Alternatively, hierarchical clustering of expression array data can reveal apparent clustering of tumors suggesting distinct sub-classifications. Investigators studying these phenomena have traditionally used clinical criteria, such as case survival, to try to validate the relevance of any postulated classification system, although validation presents its own challenges.[24] The thesis of this article is that genuine tumor sub-classifications are likely to be etiologically distinct, and that etiologic heterogeneity is, in and of itself, an important criterion for validating the relevance of any candidate sub-classification system.

How do we test for etiologic heterogeneity of tumor sub-classifications? Classically this is accomplished in case-control studies where the odds ratios of candidate or known risk factors are compared between the tumor sub-categories. Although comparison of each subtype with a common control group is frequently employed for this purpose, it has been shown that the most direct and efficient way to identify etiologic heterogeneity of individual risk factors is simply by comparing the risk profiles of the sub-classifications without the need for population controls in a case-only design.[25] This strategy is necessary to identify individual risk factors that have distinctive influences on the risk heterogeneity. However, the theory presented in this article has shown that a more global assessment of etiologic heterogeneity is possible prior to conducting case-control or case-only studies to identify the sources of the heterogeneity. That is, detailed analyses of the tumor characteristics of double primaries can, in principle, provide important insights for planning and analyzing future epidemiologic investigations. By examining in a series of double primaries the odds ratios of various candidate tumor sub-classifications based on somatic molecular profiling and/or visible pathologic or clinical characteristics, and by identifying the classification with the largest odds ratio, future epidemiologic investigations could be based on searching for risk factors that provide distinctive effects on the disease sub-categories so identified, or indeed by examining these sub-categories in separate studies.

However, this strategy does have limitations. In particular one needs studies of double primary malignancies where the tissue from both tumors is available in order that the necessary pathological classifications can be accomplished on both tumors. The examples presented of breast cancers typically involve relatively small studies in settings where the ingredients of the sub-classification (ER, PR and histology) are collected routinely. Prospective studies of new approaches based on, say genome-wide arrays, would involve logistical challenges to collect the necessary specimens in sufficient numbers and quality to accomplish the sub-classifications of the pairs of tumors. In addition to these technical and logistical challenges the proposed method can really only be contemplated for cancer sites where the occurrence of independent second primaries are relatively common. This certainly includes breast cancer and melanoma, and possibly sites such as lung and colorectal, though care would be necessary to identify genuinely independent second primaries. Studies of paired organs such as the ovary, the testicle and the kidney are also feasible in principle, though limited by the rarity of occurrence of double malignancies. These sample size limitations also affect the feasibility of studying rare tumor sub-types.

The data from studies of breast cancer show a strong concordance of double primaries on the basis of ER receptor status. This result is consistent with a growing literature of epidemiologic studies that have identified specific reproductive risk factors that differ markedly in their effect on the risks of ER+ versus ER- breast cancers.[26–28] Studies have also been conducted that point to genetic factors that distinguish ER+ from ER- tumors.[29,30] Interestingly, it has also been shown that risk factors distinguish breast cancer histologic types.[31] Investigators have also studied the etiologic heterogeneity of more refined molecular sub-types, such as those based on Her2 expression in addition to ER and PR status,[32,33] categories originally suggested by expression profiling.[34] A more detailed investigation in a large study of paired contralateral breast tumors would be necessary to determine the combination of receptor status, histology, and possibly other tumor characteristics that provide the axes on which the etiologic heterogeneity of breast cancer is best represented.

The theory on which the results are based also has technical limitations. It involves the premise that cancer occurrence is a fundamentally stochastic phenomenon that is influenced by genetic and environmental risk propensities that are unique to the individual. We must assume that this individual cancer propensity is the predominant influence on the risk of both the first cancer and the second cancer. This allows us to assume that the two occurrences are essentially experimental replicates, which in turn allows us to infer the degree of person-to-person risk variation in the population (see the Statistical Appendix for further details). Further, the assumption allows us to infer indirectly the risk covariance between the sub-classes, i.e. the degree of risk heterogeneity. This assumption is clearly not literally correct, and it could be perturbed by issues such as the differential impact of treatment for the first primary on the risk of a second primary of the different sub-types, differences in case survival of the sub-types, changing underlying risk due to the aging of the individual, and diagnostic errors, such as misclassifications of metastases as second primaries (or vice versa). Indeed, in the literature on ER status in breast cancer the authors of the studies reported in Table 2 were frequently focused on the impact of hormonal treatment on the receptor status of the second primary. Also, although it is conventional to classify contralateral breast cancers as “independent” second primaries, in fact this is not a settled issue. In studies of the clonal relatedness of contralateral breast tumors the mutational profiles of synchronous tumors often appear more similar than for metachronous tumors (see for example Imyanitov et al.35). In fact, in the studies presented in Table 2, the ER/PR profiles of synchronous cases generally exhibited greater association than for the metachronous cases presented in the table (data not shown), suggesting that a proportion of contralateral breast cancers may actually be metastases from the first primary, with this proportion being higher for more contemporaneous (i.e. synchronous) occurrences. These potential problems should dissuade us from over-interpreting the magnitude of observed concordance odds ratios. However, the thesis is that the approach is nonetheless valuable as a tool for identifying risk heterogeneity from a broad brush perspective, and for comparing and ranking classification systems on their concordances.

In summary, observation of the relative frequencies of concordances and discordances of sub-types of double primary malignancies provides a unique opportunity to gain insight into cancer epidemiology. The degree of concordance provides direct evidence of the global risk heterogeneity of the sub-categories, providing experimental validation of the etiologic relevance of the sub-classification system. The presence of strong etiologic heterogeneity points to the need for epidemiologic investigations that involve separate study of the sub-types to search for the distinct risk factors (or distinct effects of individual risk factors) that are causing the heterogeneity. Failure to conduct stratified epidemiologic studies in the presence of strong etiologic heterogeneity inevitably diminishes the sensitivity and statistical power of the study to detect important risk factors whose effects are different for the sub-types. Careful evaluation of the concordance of tumor characteristics in double primaries should be an important tool in the on-going effort to uncover the causes of cancer.

Acknowledgments

I am grateful to Malcolm Pike for insightful comments on an early draft of this article, to Monica Brown for supplying detailed data from her study using the California Cancer Registry, and to Erica Schubert for help with the literature review. This work was supported by the National Cancer Institute at the National Institutes of Health (CA124504 and CA131010).

STATISTICAL APPENDIX

Every individual in the population can be classified into one of many risk categories based on the magnitude of the risk. Let the prevalence of the ith risk category be p_i where $\sum p_i = 1$. Let the cancer risk for individuals in the ith category be r_i, where this risk is the sum of the cancer risks of the tumor sub-types, i.e. $r_i = r_{Ai} + r_{Bi}$, where r_{Ai} is the risk of a cancer of type A, and r_{Bi} is the risk of a cancer of type B. Further, let the corresponding mean risks be, respectively, μ, μ_A and μ_B where $\mu = \mu_A + \mu_B$. We can represent the relative degree to which risk varies in the population by $K^2 = v/\mu^2$, the square of the coefficient of variation of the distribution of risks in the population, where $v=\sum p_i r_i^2 - \mu^2$. Similarly, the risk variation coefficients for sub-types A and B are denoted $K_A^2 = v_A/\mu_A^2$ and $K_B^2 = v_B/\mu_B^2$, where $v_A = \sum p_i r_{Ai}^2 - \mu_A^2$ and $v_B = \sum p_i r_{Bi}^2 - \mu_B^2$. Further let c be the covariance in the risk profiles, i.e. $c = \sum p_i\, r_{Ai}\, r_{Bi} - \mu_A\, \mu_B$, and define the risk covariance coefficient as $K_C = c / \mu_A\, \mu_B$.

These stratified population risks are the fundamental forces that drive the observed occurrences of cancer types A and B in patients who experience double malignancies. Defining $E_{AA}=E(F_{AA}/N)$, $E_{AB}=E(F_{AB}/N)$, etc. as the long-run expected frequencies of these co-occurrences, we can express these as a function of the underlying risk correlation structure as follows. Consider E_{AA} first. E_{AA} represents the probability that both the first and second tumors are of sub-type A, given that a double malignancy has been observed in the patient. This can be further decomposed into the probability that the first tumor is of type A, multiplied by the probability that the second tumor is of type A, given that the first is of type A. The probability that the first tumor is of type A is simply μ_A / μ. However to calculate the probability associated with the second tumor we need to understand the concept of risk-biased sampling. Patients with tumors of type A are sampled in direct proportion to their risks of a type A tumor.36 If, say, people in risk group i have twice the risk of those in risk group j, i.e. $r_{Ai} = 2r_{Aj}$, then they are twice as likely to be sampled (i.e. to have a cancer of type A occur). In fact, more generally, risk group i will be represented among individuals with cancer of type A in proportion to $p_i\, r_{Ai}$. Thus in the risk profile for patients with A tumors, the population frequencies p_1, p_2, p_3, … are replaced by q_{A1}, q_{A2}, q_{A3},… where $q_{Ai} = p_i\, r_{Ai} / \sum p_i\, r_{Ai}$. It follows that among patients with double primaries in which the first tumor was of type A the probability that the second tumor is also of type A is given by the following:-

$$\Pr(\text{second tumor is A given first tumor is A}) = \frac{\sum q_{Ai} r_{Ai}}{\sum q_{Ai} r_{Ai} + \sum q_{Ai} r_{Bi}}$$
$$= \frac{\sum p_i r_{Ai}^2}{\sum p_i r_{Ai}^2 + \sum p_i r_{Ai} r_{Bi}}.$$

Consequently

$$E_{AA} = \frac{\mu_A}{\mu} \frac{\sum p_i r_{Ai}^2}{\sum p_i r_{Ai}^2 + \sum p_i r_{Ai} r_{Bi}}$$

and by a similar derivation

$$E_{AB} = \frac{\mu_A}{\mu} \frac{\sum p_i r_{Ai} r_{Bi}}{\sum p_i r_{Ai}^2 + \sum p_i r_{Ai} r_{Bi}}$$

$$E_{BA}=\frac{\mu_B}{\mu}\frac{\sum p_i r_{Ai} r_{Bi}}{\sum p_i r_{Bi}^2+\sum p_i r_{Ai} r_{Bi}}$$

and

$$E_{BB}=\frac{\mu_B}{\mu}\frac{\sum p_i r_{Bi}^2}{\sum p_i r_{Bi}^2+\sum p_i r_{Ai} r_{Bi}}.$$

From this it is easily shown that

$$\psi=\frac{E_{AA}E_{BB}}{E_{AB}E_{BA}}=\frac{\left(1+K_A^2\right)\left(1+K_B^2\right)}{\left(1+K_C\right)^2}.$$

It is important to recognize that the preceding mathematical structure is constructed for conceptual clarity, but clearly contains some important assumptions and approximations. The key assumption is that two cancers that occur in the same patient are experimental replicates. To make this assumption we must be confident that the tumors are biologically independent. This would not be the case if one of the tumors is actually a metastasis of the first tumor, mis-diagnosed as a second primary. There is a considerable recent literature on this evolving topic of investigation, but current thinking is that for some cancers, notably contralateral breast cancer and melanoma, we can be confident that most diagnosed second primaries are biologically independent, while for others metastases may frequently be misdiagnosed as second primaries.[35,37–40] Clearly frequent misdiagnoses of this nature would inevitably inflate the apparent association of tumor sub-types. The notion of experimental replication also requires that we view the probability of occurrence of a first cancer in an individual as being the same as the probability of a second cancer, given the occurrence of the first. The idea here is that the cancer risk of any individual is approximately constant over the period in which the two cancers occur. This assumption is clearly not literally true, since cancer risk changes with age, and the second cancer necessarily occurs at a later age. However, the relatively few years that usually elapse between most observed double primaries suggests that the influence of age on this phenomenon is probably minor. A more serious concern is that treatment for the first primary, especially systemic medical treatment as opposed to surgery or radiotherapy, may alter the risk profile for the subsequent cancer.[41] Finally we assume population-based incidence sampling of patients with a second cancer. In this way we can relate the occurrence frequencies directly with the cancer risks in the underlying population. Because of these issues we must view the resulting analysis as providing broad, overarching inferences that can be useful for planning research strategies, rather than an analysis that provides precise estimates of effect.

REFERENCES

1. Harris TJR, McCormick F. The molecular pathology of cancer. Nat Rev Clin Oncol. 2010; 7:251–265. [PubMed: 20351699]
2. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, Alexe G, Lawrence M, et al. Integrated genomic analysis identifies clinically relevant

subtypes of glioblastoma characterized by abnormalities in *PDGFRA*, *IDH1*, *EGFR*, and *NF1*. Cancer Cell. 2010; 17:98–110. [PubMed: 20129251]

3. Barretina J, Taylor BS, Banerji S, Ramos AH, Lagos-Quintana M, Decarolis PL, Shah K, Socci ND, Weir BA, Ho A, Chiang DY, Reva B, et al. Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. Nat Genet. 2010; 42:715–721. [PubMed: 20601955]
4. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, Arora VK, Kaushik P, Cerami E, Reva B, Antipin Y, Mitsiades N, et al. Integrative genomic profiling of human prostate cancer. Cancer Cell. 2010; 18:11–22. [PubMed: 20579941]
5. Troester MA, Swift-Scanlan T. Challenges in studying the etiology of breast cancer subtypes. Breast Cancer Res. 2009; 11:104. [PubMed: 19635173]
6. Travis LB, Rabkin CS, Brown LM, Allan JM, Alter BP, Ambrosone CB, Begg CB, Caporaso N, Chanock S, DeMichele A, Figg WD, Gospodarowicz MK, et al. Cancer survivorship—genetic susceptibility and second primary cancers: research strategies and recommendations. J Natl Cancer Inst. 2006; 98:15–25. [PubMed: 16391368]
7. Kuligina E, Reiner A, Imyanitov EN, Begg CB. Evaluating cancer epidemiologic risk factors using multiple primary malignancies. Epidemiology. 2010; 21:366–372. [PubMed: 20299982]
8. Begg CB, Berwick M. A note on the estimation of relative risks of rare genetic susceptibility markers. Cancer Epidemiol Biomarkers Prev. 1997; 6:99–103. [PubMed: 9037560]
9. Imyanitov EN, Cornelisse CJ, Devilee P. Searching for susceptibility alleles: emphasis on bilateral breast cancer. Int J Cancer. 2007; 121:921–923. [PubMed: 17444500]
10. Peto J, Mack TM. High constant incidence in twins and other relatives of women with breast cancer. Nat Genet. 2000; 26:411–414. [PubMed: 11101836]
11. Arpino G, Weiss HL, Clark GM, Hilsenbeck SG, Osborne CK. Hormone receptor status of a contralateral breast cancer is independent of the receptor status of the first primary in patients not receiving adjuvant tamoxifen. J Clin Oncol. 2005; 23:4687–4694. [PubMed: 15837971]
12. Bachleitner-Hofmann T, Pichler-Gebhard B, Rudas M, Gnant M, Taucher S, Kandioler D, Janschek E, Dubsky P, Roka S, Sporn E, Jakesz R. Pattern of hormone receptor status of secondary contralateral breast cancers in patients receiving adjuvant tamoxifen. Clin Cancer Res. 2002; 8:3427–3432. [PubMed: 12429630]
13. Gong SJ, Rha SY, Jeung HC, Roh JK, Yang WI, Chung HC. Bilateral breast cancer: differential diagnosis using histological and biological parameters. Jpn J Clin Oncol. 2007; 37:487–492. [PubMed: 17673471]
14. Hahnel R, Twaddle E. The relationship between estrogen receptors in primary and secondary breast carcinomas and in sequential primary breast carcinomas. Breast Cancer Res Treat. 1985; 5:155–163. [PubMed: 4016281]
15. Holdaway IM, Mason BH, Bennett RC, Alexander AI, Hahnel R, Kiang DT. Estrogen receptors in bilateral breast cancer. Cancer. 1988; 62:109–113. [PubMed: 3383109]
16. Idvall I, Ringberg A, Anderson H, Akerman M, Fernö M. Histopathological and cell biological characteristics of ductal carcinoma in situ (DCIS) of the breast-a comparison between the primary DCIS and subsequent ipsilateral and contralateral tumours. Breast. 2005; 14:290–297. [PubMed: 16085235]
17. Kollias J, Pinder SE, Denley HE, Ellis IO, Wencyk P, Bell JA, Elston CW, Blamey RW. Phenotypic similarities in bilateral breast cancer. Breast Cancer Res Treat. 2004; 85:255–261. [PubMed: 15111764]
18. Stark A, Lu M, Mackowiak P, Linden M. Concordance of the hormone receptors and correlation of HER-2/neu overexpression of the metachronous cancers of contralateral breasts. Breast J. 2005; 11:183–187. [PubMed: 15871703]
19. Swain SM, Wilson JW, Mamounas EP, Bryant J, Wickerham DL, Fisher B, Paik S, Wolmark N. Estrogen receptor status of primary breast cancer is predictive of estrogen receptor status of contralateral breast cancer. J Natl Cancer Inst. 2004; 96:516–523. [PubMed: 15069113]
20. Weitzel JN, Robson M, Pasini B, Manoukian S, Stoppa-Lyonnet D, Lynch HT, McLennan J, Foulkes WD, Wagner T, Tung N, Ghadirian P, Olopade O, et al. A comparison of bilateral breast cancers in BRCA carriers. Cancer Epidemiol Biomarkers Prev. 2005; 14:1534–1538. [PubMed: 15941968]

21. Brown M, Bauer K, Pare M. Tumor marker phenotype concordance in second primary breast cancer, California, 1999–2004. Breast Cancer Res Treat. 2010; 120:217–227. [PubMed: 19629680]

22. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltnane JM, Hurt EM, Zhao H, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med. 2002; 346:1937–1947. [PubMed: 12075054]

23. Yeoh EJ, Ross ME, Shurtleff SA, Williams WK, Patel D, Mahfouz R, Behm FG, Raimondi SC, Relling MV, Patel A, Cheng C, Campana D, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer Cell. 2002; 1:133–143. [PubMed: 12086872]

24. Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. Lancet. 2005; 365:488–492. [PubMed: 15705458]

25. Begg CB, Zhang ZF. Statistical analysis of molecular epidemiology studies employing caseseries. Cancer Epidemiol Biomarkers Prev. 1994; 3:173–175. [PubMed: 8049640]

26. Althuis MD, Fergenbaum JH, Garcia-Closas M, Brinton LA, Madigan MP, Sherman ME. Etiology of hormone receptor-defined breast cancer: a systematic review of the literature. Cancer Epidemiol Biomarkers Prev. 2004; 13:1558–1568. [PubMed: 15466970]

27. Chen WY, Colditz GA. Risk factors and hormone-receptor status: epidemiology, risk-prediction models and treatment implications for breast cancer. Nat Clin Pract Oncol. 2007; 4:415–423. [PubMed: 17597706]

28. Ma H, Bernstein L, Pike MC, Ursin G. Reproductive factors and breast cancer risk according to joint estrogen and progesterone receptor status: a meta-analysis of epidemiologic studies. Breast Cancer Res. 2006; 8(4):R43. [PubMed: 16859501]

29. Garcia-Closas M, Hall P, Nevanlinna H, Pooley K, Morrison J, Richesson DA, Bojesen SE, Nordestgaard BG, Axelsson CK, Arias JI, Milne RL, Ribas G, et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. PLoS Genet. 2008; 4(4):e1000054. [PubMed: 18437204]

30. Garcia-Closas M, Chanock S. Genetic susceptibility loci for breast cancer by estrogen receptor status. Clin Cancer Res. 2008; 14:8000–8009. [PubMed: 19088016]

31. Reeves GK, Pirie K, Green J, Bull D, Beral V. Reproductive factors and specific histological types of breast cancer: prospective study and meta-analysis. Br J Cancer. 2009; 100:538–544. [PubMed: 19190634]

32. Kwan ML, Kushi LH, Weltzien E, Maring B, Kutner SE, Fulton RS, Lee MM, Ambrosone CB, Caan BJ. Epidemiology of breast cancer subtypes in two prospective cohort studies of breast cancer survivors. Breast Cancer Res. 2009; 11:R31. [PubMed: 19463150]

33. Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, Hewitt SM, Anderson WF, Szeszenia-Dabrowska N, Bardin-Mikolajczak A, Zatonski W, Cartun R, et al. Differences in risk factors for breast cancer molecular subtypes in a population-based study. Cancer Epidemiol Biomarkers Prev. 2007; 16:439–443. [PubMed: 17372238]

34. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, et al. Molecular portraits of human breast tumours. Nature. 2000; 406:747–752. [PubMed: 10963602]

35. Imyanitov EN, Suspitsin EN, Grigoriev MY, Togo AV, Kuligina ESh, Belogubova EV, Pozharisski KM, Turkevich EA, Rodriquez C, Cornelisse CJ, Hanson KP, Theillet C. Concordance of allelic imbalance profiles in synchronous and metachronous bilateral breast carcinomas. Int J Cancer. 2002; 100:557–564. [PubMed: 12124805]

36. Begg CB. On the use of familial aggregation in population-based case probands for calculating penetrance. J Natl Cancer Inst. 2002; 94:1221–1226. [PubMed: 12189225]

37. Ha PK, Califano JA. The molecular biology of mucosal field cancerization of the head and neck. Crit Rev Oral Bio Med. 2003; 14:363–369. [PubMed: 14530304]

38. Hafner C, Knuechel R, Stoehr R, Hartmann A. Clonality of multifocal urothelial carcinomas: 10 years of molecular genetic studies. Int J Cancer. 2002; 101:1–6. [PubMed: 12209580]

39. Orlow I, Tommasi DV, Bloom B, Ostrovnaya I, Cotignola J, Mujumdar U, Busam KJ, Jungbluth AA, Scolyer RA, Thompson JF, Armstrong BK, Berwick M, et al. Evaluation of the clonal origin of multiple primary melanomas using molecular profiling. J Invest Dermatol. 2009; 129:1972–1982. [PubMed: 19282844]

40. Girard N, Ostrovnaya I, Lau C, Park B, Ladanyi M, Finley D, Deshpande C, Rusch V, Orlow I, Travis WD, Pao W, Begg CB. Genomic and mutational profiling to assess clonal relationships between multiple non-small cell lung cancers. Clin Cancer Res. 2009; 15:5184–5190. [PubMed: 19671847]

41. Neugut, AI.; Meadows, AT.; Robinson, E., editors. Multiple Primary Cancers. Philadelphia: Lippincott Williams & Wilkins; 1999.

Table 1

Basic Data Structure and Terms[1]

		2nd Tumor	
		Sub-Type A	**Sub-Type B**
1st Tumor	Sub-Type A	F_{AA}	F_{AB}
	Sub-Type B	F_{BA}	F_{BB}

[1] The entries in the table, F_{AA}, F_{AB}, F_{BA} and F_{BB}, represent the frequencies of patients' tumors cross-classified by tumor sub-type among a total sample of N patients.

Table 2

Concordances of Tumor Characteristics in Contralateral Breast Primaries

	1st Primary/2nd Primary[a]				
Study	**ER+/ER+**	**ER+/ER−**	**ER−/ER+**	**ER−/ER−**	**Odds Ratio[c]**
Arpino et al.11,[b]	115	38	26	14	1.6
Bachleitner-Hoffman12,[d]	22	3	5	5	7.3
Brown et al.21,[e]	279	53	40	70	9.2
Gong et al.13	14	8	5	10	3.5
Hahnel et al.14	11	1	5	3	6.6
Holdaway et al.15	14	4	7	10	5.0
Idvall et al.16	4	1	5	2	1.6
Kollias et al.17	26	8	6	6	3.3
Stark et al.18	28	19	3	11	5.4
Swain et al.19	53	21	10	26	6.6
Weitzel et al.20	49	21	16	105	15.3
Totals	**615**	**177**	**128**	**262**	**5.2 [3.8–7.2]**[f]
	PR+/PR+	**PR+/PR−**	**PR−/PR+**	**PR−/PR−**	
Arpino et al.11,[b]	57	49	35	37	1.2
Bachleitner-Hoffman12,[d]	8	4	12	11	1.8
Brown et al.21,[e]	183	108	55	96	3.0
Gong et al.13	7	9	4	19	3.7
Idvall et al.16	4	1	4	3	3.0
Kollias et al.17	16	12	7	11	2.1
Stark et al.18	23	18	12	8	0.9
Totals	**298**	**201**	**129**	**185**	**2.1 [1.6–2.9]**

[a] The numbers in the table represent frequencies of patients with the designated tumor characteristics. Except where indicated, data restricted to metachronous occurrences of second primaries.

[b] Includes a combination of synchronous and metachronous cases.

[c] Bolded numbers are summary odds ratios and 95% confidence intervals. Tests for heterogeneity were p=0.03 for ER (excluding the study by Weitzel et al.) and p=0.15 for PR.

[d] Restricted to Tamoxifen-treated patients

[e] Raw data not presented in sufficient detail in the reference: the data in the table were kindly supplied by Monica Brown (California Cancer Registry)

[f] The summary odds ratio excludes the study by Weitzel that was restricted to carriers of BRCA1 or BRCA2 mutations; the test for heterogeneity leads to p=0.003 when this study is included.