HHS Public Access

Author manuscript

Cancer Epidemiol Biomarkers Prev. Author manuscript; available in PMC 2017 August 08.

Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2010 November ; 19(11): 2710–2714. doi: 10.1158/1055-9965.EPI-10-0742.

What can we Learn about Disease Etiology from Case-Case Analyses? Lessons from Breast Cancer[1]

María Elena Martínez[1,2], **Giovanna Cruz**[1,2], **Abenaa Brewster**[3], **Melissa Bondy**[3], and **Patricia A. Thompson**[1,2]

[1]Arizona Cancer Center, University of Arizona, Tucson, Arizona

[2]Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, Arizona

[3]University of Texas M.D. Anderson Cancer Center, Houston, Texas

Introduction

In this commentary, we discuss the challenges and opportunities for epidemiologic studies in evaluating breast cancer as a set of discrete diseases. We show examples of the strengths of the case-only design in assessing the relative correlation of established risk factors and the different subtypes. We argue for the use of the case-only study design as an important initial step in understanding the extent of etiologic heterogeneity between tumor subtypes.

Breast Tumor Heterogeneity and Etiology

Unlike cardiovascular disease, where distinct disease subtypes and etiologic factors are recognized (e.g., ischemic vs. hemorrhagic stroke), cancer is in its infancy in regard to an understanding of distinct causal pathways. However, breast cancer is perhaps in an adolescent stage given the important recent advances in the genomic characterization of breast tumors. These advances have shown that breast cancers segregate reproducibly into discrete groups that differ in their clinical behaviors (1–3). The molecular subtypes identified by gene expression profiles align, expectedly, in the first order of clustering algorithms along the estrogen receptor (ER) status of the tumor (1). Motivation for second order classification and characterization stems from an important and overarching interest to understand the clinical significance of the second and third order heterogeneity in terms of patient outcomes and drug discovery efforts. This latter effort has led to an evolving set of classifications of breast cancers that encompass at least four *clinically* relevant and highly reproducible subclasses: luminal A, luminal B, *ERBB2* amplified (HER2+), and basaloid type tumors. Thus, from a patient outcomes perspective, it is clear that a second and perhaps, third order heterogeneity exists among breast tumors that is clinically meaningful. However, for epidemiology, where the interest is in the *etiologic perspective,* it is unclear how much

[1]Grant Support: Work was supported by the Avon Foundation, a supplement to the Arizona Cancer Center Core Grant from the National Cancer Institute (CA-023074-2953), a supplement to the M.D. Anderson Cancer Center SPORE in Breast Cancer (P50 CA116199-02S1), and by a grant from Susan G. Komen for the Cure (KG090934).

Corresponding Author: María Elena Martínez, Ph.D., University of Arizona, Arizona Cancer Center, PO Box 245024, Phone: 520-626-8130, Fax: 520-626-9275, emartinez@azcc.arizona.edu.

this heterogeneity results from clonal expansion and evolution of a common progenitor cell or simply the consequence of carcinogenesis in separate cell lineages for which a distinct constellation of causal factors exist.

The view of breast cancer, not as a set of stochastic molecular events, but as some finite set of separate tumorigenesis events, is not entirely new to genomics. In 1995, Potter *et al.,* provided some of the earliest evidence supporting etiologic heterogeneity among breast cancers showing an inverse association between parity and ER/progesterone receptor (PR) positive tumors that was not present for ER/PR negative tumors (4). A decade later, the Nurses' Health Study reported that parity and early age of first pregnancy were inversely associated with ER^+/PR^+, but not with ER^-/PR^- tumors (5); these data were replicated in a later meta-analysis (6). More recent studies support the presence of distinct reproductive and genetic risk factor profiles among the tumor subtypes that largely divide on hormone receptor status (7–9). Additional evidence in support of distinct disease subtypes derives from studies of tumors arising in *BRCA1* and *BRCA2* germline mutation carriers. *BRCA1* carriers develop basaloid tumors (10, 11), for which early first pregnancy acts as a risk factor (12). In contrast, *BRCA2* carriers develop luminal tumors, for which early first pregnancy is protective (12–14). This opposing effect of age of first pregnancy and tumor type-specific risk in the *BRCA* cases provides some of the strongest evidence that breast cancer subtypes can and likely do arise through distinct causal pathways.

Studying the Epidemiology of Breast Tumor Subtypes

Recent developments in breast tumor classification based on molecular characteristics pose significant methodological challenges for conventional epidemiologic studies of breast cancer etiology and risk assessment. First, there is the lack of consensus on disease classification and subtype definitions. Ignoring the lack of consensus, routinely reported tumor markers such as ER and PR status are insufficient to categorize cases into the newly described molecular-based subtypes. Therefore, tissue collection from cases is necessary to obtain additional gene and protein measures to derive the classification. However, tissue collection presents its own set of challenges for population-based epidemiologic studies including costs, issues of ownership and consent, and increasingly smaller tumors and biopsy-based sampling. Second, depending on the number of subtypes and their prevalence, traditional cohort and case-control studies face problems related to statistical power; the result is study findings that are inconsistent, imprecise, and ultimately, uninformative. A few recent studies have attempted to begin to address these challenges by pooling data from numerous studies and stratifying on ER and PR status and other clinical features (15, 16). These pooled studies have mainly focused on genetic susceptibility loci as the main risk factors of interest. Integration of epidemiological exposure variables has proven more challenging for the pooling approach because of non-standard collection of information on exposures. Additional challenges specific to case-control studies include appropriate control selection and low response rates among controls. This latter issue is especially true for studies of racial/ethnic minorities and other disadvantaged populations, where we now have evidence of different disease patterns for the breast cancer subsets (9). In these populations, it is also essential to consider matching on race/ethnicity or to employ family-based studies to minimize population stratification.

Case-Case Analysis of Breast Cancer Etiology

In a case-only study design, information is obtained only from cases of a particular disease (i.e., breast cancer), with no information obtained from individuals without the disease. In genetic epidemiology, the case-only design has been used to study the effects of the interaction between environmental factors and genetic variants (17). Assuming independence between the genetic and environmental factor, a differential association with the environmental risk factor of interest by genotype is interpreted as supportive of etiological heterogeneity (17, 18). In addition, as early as 1984, Prentice et al. (19) introduced the concept of using a case-only study for identifying disease risk factors. More recently, case-case analyses have been used as an exploratory tool to assess etiological heterogeneity in the context of breast cancer; here, the most common breast tumor subtype (luminal A) is used as the reference group against which the other subtypes are compared for their odds of having a given exposure. The rationale is that a difference in the relationship between the exposure supports, but does not necessarily prove, the hypothesis of distinct effects on the discrete disease subsets.

Comparison of Case-case and Case-control Analyses

To illustrate our argument for the use of the case-only analysis, we use data from two published studies that have explored breast cancer etiologic heterogeneity. The first is from the Carolina Breast Cancer Study (CBCS), which is a population-based case-control study of African-American and white women in North Carolina (9). The second is a pooled analysis of data from five individual studies used to assess genetic susceptibility loci as risk factors for breast cancer (16).

The CBCS included 1,424 cases of invasive and *in situ* breast cancer and 2,022 disease free controls. The study examined associations for a variety of traditional breast cancer risk factors in relation to five immunohistochemical (IHC) tumor marker defined subtypes. This study setting provided a unique opportunity to conduct a parallel comparison of the results from a case-case and case-control analysis for a variety of reproductive and lifestyle factors. When assessing age at menarche (< 13 vs. ≥ 13 years), the case-case odds ratio (OR) was 1.3 for basal-like cases (i.e., tumors that are ER−, PR−, HER2−, HER1+ and/or CK5/6+) compared to luminal A cases (i.e., tumors that are ER+ and/or PR+ and HER2−). Here, the case-case OR represents the ratio of the ORs for the different subtypes and shows that cases with basal-like tumors are more likely to report an earlier age at menarche when compared to cases with luminal A like tumors. In the case-control analysis, where disease-free controls are used as the comparison group, cases with luminal A tumors had a 1.1 odds of reporting an earlier age of menarche, whereas the corresponding odds for cases with basal-like tumors was 1.4. Based on the difference in magnitude and direction of the association in the case-control setting (luminal A OR 1.1 vs. basal-like OR 1.4), we interpret this relationship to mean that cases with basal-like tumors are more likely to have an earlier age of menarche than cases with luminal A tumors. These findings are consistent with the interpretation from the case-only analysis. Perhaps a more interesting comparison between the two designs is related to parity, where there is prior evidence for a differential effect of parity by tumor subtype (5, 6, 20). The case-control analysis shows that compared to nulliparous controls,

cases with luminal A tumors had a lower odds of having 1, 2 or 3 or more births (OR of 0.7 for all three categories). Conversely, cases with basal-like tumors had an increased odds of reporting higher parity compared to nulliparous controls. Furthermore, for cases with basal-like tumors, the association strengthened with additional births; OR 1.7, 1.8, and 1.9 for 1, 2, and 3 or more births, respectively. In the case-case analysis, basal-like tumors were more likely to occur among parous women compared to luminal A tumors (OR=1.6 for 1–2 children and 1.7 for 3+ children). In addition to these examples, concordant results were shown for breastfeeding, waist-to-hip ratio, and others between the case-case and case-control analyses in the CBCS. Thus, although these empirical results could be subject to bias, the comparisons illustrate that the information gained for the association between established breast cancer risk factors and tumor subtype is generally comparable using the two study design approaches.

This conclusion is further strengthened by recent analyses of the genetic studies in breast cancer. For example, as in the case of the CBCS, Stacey et al. (16) conducted case-control and case-case analyses of genetic variants on chromosome 5p12 in relation to breast cancer risk. Unlike the CBCS, this study did not have data on the intrinsic tumor subtypes; therefore, susceptibility was only assessed by ER status. In the case-control analysis, the OR for the presence of the T allele in the SNP rs4415084 and breast cancer risk was 1.16; however, stratification by ER status showed no association for ER negative tumors vs. controls (OR=0.98) and a positive association for ER positive tumors (OR=1.23). The case-case OR for ER positive vs. ER negative tumors was 1.25, which is merely an estimate of the ratio of the ORs (1.23/0.98). In this scenario, both ORs are essentially identical given that no association is apparent between the genetic locus and ER negative tumors. It is also important to point out that this study comprised a large population (5,028 cases and 32,090 controls), providing adequate statistical power to uncover this degree of heterogeneity.

Interpretation of Case-only Odds Ratios and Etiologic Heterogeneity

In Figure 1, we illustrate the findings of case-control and case-case analyses of breast cancer in the presence and absence of etiologic heterogeneity. In the presence of heterogeneity (left panel), parity, for example, would be positively associated with ER− tumors and negatively associated with ER+ tumors (5, 6, 20). In this setting of etiologic heterogeneity, the case-control OR will reflect the differential direction of the effect and the case-case OR, which is an estimate of the ratio of the case-control ORs for ER subtypes, will have a value greater than 1 when comparing ER− vs. ER+ disease subtypes. For the absence of heterogeneity (right panel), we use the genetic locus on *MAP2K1,* which has been reported to be positively associated with risk of both ER+ and ER− tumor subtypes (21). In this case, the risk factor is positively associated with risk of both tumor subtypes; the case-control ORs will show a value greater than 1 whereas the case-case OR will be null or close to this value. In this scenario, the case-only study is interpreted as a failure to demonstrate etiologic heterogeneity for this risk factor between the tumor subtypes.

Given the proposed disease heterogeneity observed in breast cancer, future large epidemiologic studies will be helpful in identifying etiologic heterogeneity for the established, traditional risk factors by disease subtype. However, it must be noted that our

knowledge is still incomplete given that to date, sample sizes for the less common tumor subtypes have been limited. Furthermore, as noted earlier, the precise, clinically significant tumor subtypes are still unknown and this continues to be an active area of research. Though highly speculative, the presence of common chromosomal aberrations in basal-like, HER2 and luminal B type tumors that are distinct from those arising in luminal A type tumors has led to the suggestion that etiology may be limited to two distinct cancer progenitors again largely splitting on the estrogen receptor status (22). These studies highlight the need for further integration of molecular classification to refine our definition of the disease outcome in epidemiologic studies of breast cancer.

Limitations of Case-Case Analyses

In spite of its strengths, the case-only study design has obvious limitations. Importantly, a study that does not include a disease-free population does not provide a traditional risk ratio and may not provide a valid estimate of the association between a risk factor and disease. The resulting OR can only be interpreted as the ratio of the odds of exposure for a given subtype (i.e., basal-like) in reference to another (i.e., luminal A). Thus, the case-only OR can never be interpreted as a measure of risk for the specific subtype. Furthermore, the magnitude of the association is not the magnitude of risk, but rather an indicator of the general direction of the correlation between risk factor and subtype. For example, in the CBCS, the case-case OR for basal-like compared to luminal A tumors for women who reported having 3 or more children under-estimates the true association given that the corresponding case-control OR analysis shows a higher estimate of the effect (1.9 in the case-control analysis vs. 1.7 in the case-only analysis). Arguably, the magnitude of the error in the estimate will depend on the risk factor's association with the case group that is used as the comparison (i.e., luminal A or ER subtype). In the study by Stacey et al., (16) discussed above, the case-case and case-control ORs were essentially identical given that the effect of the genetic locus was nearly entirely confined to ER positive tumors.

Utility of the Case-Only Study Design for Studies of Breast Cancer

We propose that the case-only approach can be particularly useful to uncover the differences or similarities in association between a risk factor and a tumor subtype. This study design is an important and efficient initial step in defining risk factor profiles for each subtype and providing first level evidence for etiological heterogeneity. In fact, published studies that have incorporated tumor marker information, such as ER and PR status, have begun to uncover etiologic heterogeneity for several traditional breast cancer risk factors and disease subtype as proof of principle of the value of this strategy and the need to consider the adverse effects of lumping all breast cancers into a single outcome (5, 8, 9, 20).

The problem of low sample size and statistical power to detect etiologic heterogeneity will apply to even reasonably large studies such as the CBCS. For example, in the CBCS, while the prevalence of the basal-like tumors is approximately 16%, that for HER2+/ER− tumors is a modest 8%. As also noted earlier, it must be emphasized that the exact number of etiologically distinct breast tumor subtypes is currently unknown. Thus, case-only studies can be a useful tool for providing additional clues for collapsing groups on etiologic

similarity to guide stratification and identify potential subtype specific risk factors, as originally proposed by Prentice et al. (19).

We argue that in order to continue to move the breast cancer field forward in a rapid fashion integrating new knowledge on tumor heterogeneity, epidemiologic studies need to include tissue collection in order to integrate information on tumor markers beyond those that are traditionally available in population-based tumor registries. However, tissue collection can be problematic and poses a universal challenge to the field. One solution is to obtain tissue from cases where relatively high retrieval rates are feasible, such as in a clinic-based sample. Recognizing the interpretation limitations of this approach for the general population, clinic based sampling could be a valuable and unique strategy to investigate the etiologic correlates of tumor subtypes to advance our understanding of etiologic heterogeneity.

Conclusion

The case-only design is a useful tool in the process of building a risk factor profile by identifying correlations between risk factor and disease subtype. This utility can prove beneficial for all diseases where considerable heterogeneity is plausible. Arguably, results from case-case analyses must be interpreted with caution and validated in traditional case-control and cohort studies in order to assess risk and estimate the exact magnitude of the effect. If indeed the subtypes represent distinct disease-types, knowledge gained from the case only setting is likely to derive important information on tumor subtype specific risk factors for tailoring and testing risk prediction models that could ultimately inform risk reduction strategies.

References

1. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000; 406:747–52. [PubMed: 10963602]
2. Sorlie T. Molecular portraits of breast cancer: tumour subtypes as distinct disease entities. Eur J Cancer. 2004; 40:2667–75. [PubMed: 15571950]
3. Sorlie T, Wang Y, Xiao C, Johnsen H, Naume B, Samaha RR, et al. Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms. BMC Genomics. 2006; 7:127. [PubMed: 16729877]
4. Potter JD, Cerhan JR, Sellers TA, McGovern PG, Drinkard C, Kushi LR, et al. Progesterone and estrogen receptors and mammary neoplasia in the Iowa Women's Health Study: how many kinds of breast cancer are there? Cancer Epidemiol Biomarkers Prev. 1995; 4:319–26. [PubMed: 7655325]
5. Colditz GA, Rosner BA, Chen WY, Holmes MD, Hankinson SE. Risk factors for breast cancer according to estrogen and progesterone receptor status. J Natl Cancer Inst. 2004; 96:218–28. [PubMed: 14759989]
6. 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 epidemiological studies. Breast Cancer Res. 2006; 8:R43. [PubMed: 16859501]
7. Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, et al. Differences in risk factors for breast cancer molecular subtypes in a population-based study. Cancer Epidemiol Biomarkers Prev. 2007; 16:439–43. [PubMed: 17372238]
8. Yang XR, Pfeiffer RM, Garcia-Closas M, Rimm DL, Lissowska J, Brinton LA, et al. Hormonal markers in breast cancer: coexpression, relationship with pathologic characteristics, and risk factor associations in a population-based study. Cancer Res. 2007; 67:10608–17. [PubMed: 17968031]

9. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, et al. Epidemiology of basal-like breast cancer. Breast Cancer Res Treat. 2008; 109:123–39. [PubMed: 17578664]
10. Schneider BP, Winer EP, Foulkes WD, Garber J, Perou CM, Richardson A, et al. Triple-negative breast cancer: risk factors to potential targets. Clin Cancer Res. 2008; 14:8010–8. [PubMed: 19088017]
11. Anders CK, Carey LA. Biology, metastatic patterns, and treatment of patients with triple-negative breast cancer. Clin Breast Cancer. 2009; 9(Suppl 2):S73–81. [PubMed: 19596646]
12. Kotsopoulos J, Lubinski J, Lynch H, Klijn J, Ghadirian P, Neuhausen S, et al. Age at first birth and the risk of breast cancer in BRCA1 and BRCA2 mutation carriers. Breast Cancer Research and Treatment. 2007; 105:221–8. [PubMed: 17245541]
13. Antoniou AC, Rookus M, Andrieu N, Brohet R, Chang-Claude J, Peock S, et al. Reproductive and hormonal factors, and ovarian cancer risk for BRCA1 and BRCA2 mutation carriers: results from the International BRCA1/2 Carrier Cohort Study. Cancer Epidemiol Biomarkers Prev. 2009; 18:601–10. [PubMed: 19190154]
14. Chang-Claude J, Andrieu N, Rookus M, Brohet R, Antoniou AC, Peock S, et al. Age at menarche and menopause and breast cancer risk in the International BRCA1/2 Carrier Cohort Study. Cancer Epidemiol Biomarkers Prev. 2007; 16:740–6. [PubMed: 17416765]
15. Garcia-Closas M, Hall P, Nevanlinna H, Pooley K, Morrison J, Richesson DA, et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. PLoS Genet. 2008; 4:e1000054. [PubMed: 18437204]
16. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet. 2007; 39:865–9. [PubMed: 17529974]
17. Begg CB, Zhang ZF. Statistical analysis of molecular epidemiology studies employing case-series. Cancer Epidemiol Biomarkers Prev. 1994; 3:173–5. [PubMed: 8049640]
18. Khoury MJ, Flanders WD. Nontraditional Epidemiologic Approaches in the Analysis of Gene Environment Interaction: Case-Control Studies with No Controls! Am J Epidemiol. 1996; 144:207–13. [PubMed: 8686689]
19. Prentice RL, Vollmer WM, Kalbfleisch JD. On the use of case series to identify disease risk factors. Biometrics. 1984; 40:445–58. [PubMed: 6487728]
20. Ma H, Henderson K, Sullivan-Halley J, Duan L, Marshall S, Ursin G, et al. Pregnancy-related factors and the risk of breast carcinoma in situ and invasive breast cancer among postmenopausal women in the California Teachers Study cohort. Breast Cancer Research. 2010; 12:R35. [PubMed: 20565829]
21. Garcia-Closas M, Chanock S. Genetic susceptibility loci for breast cancer by estrogen receptor status. Clin Cancer Res. 2008; 14:8000–9. [PubMed: 19088016]
22. Melchor L, Benitez J. An integrative hypothesis about the origin and development of sporadic and familial breast cancer subtypes. Carcinogenesis. 2008; 29:1475–82. [PubMed: 18596026]

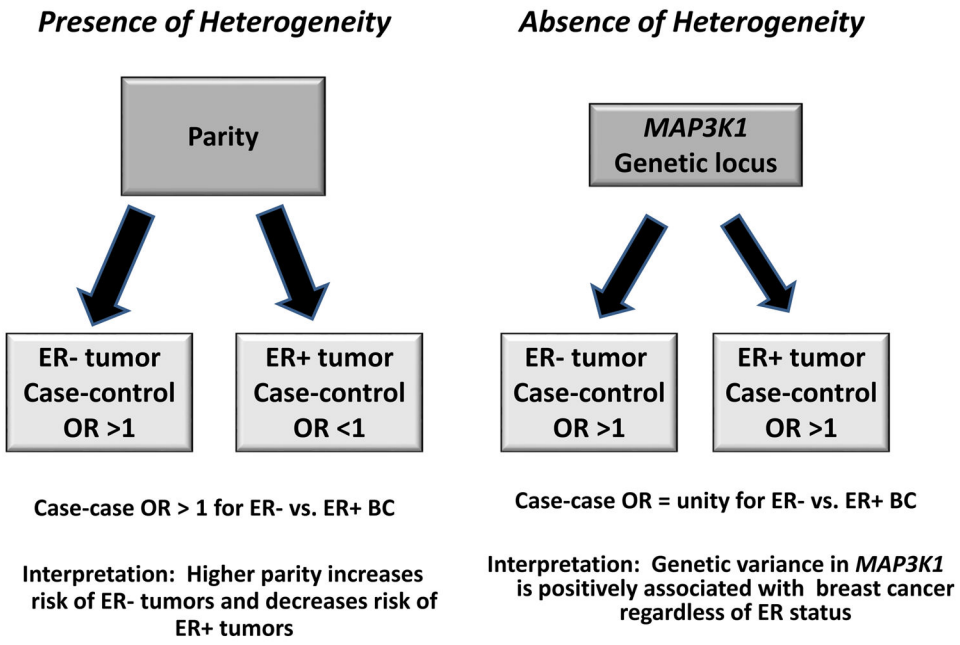

Figure 1.

Risk factor associations in the presence and absence of etiologic heterogeneity in breast cancer case-control and case-case analyses.