



Published in final edited form as:

Epidemiology. 2010 May ; 21(3): 366–372. doi:10.1097/EDE.0b013e3181cc8871.

EVALUATING CANCER EPIDEMIOLOGIC RISK FACTORS USING MULTIPLE PRIMARY MALIGNANCIES

Ekatherina Kuligina¹, Anne Reiner², Evgeny N. Imyanitov¹, and Colin B. Begg²

¹ N. N. Petrov Institute of Oncology, 68 Leningradskaya Street, Pesochny 2, 197758 St. Petersburg, Russian Federation

² Memorial Sloan-Kettering Cancer Center, 307 East 63rd Street, New York, NY 10065, USA

Abstract

Background—Several authors have proposed the use of patients with double primary malignancies affecting the same or contralateral organ as a genetically enriched resource of cases for epidemiologic case-control studies in cancer. Such an approach is based on the assumption that the factors that increase the risk of a second primary are the same ones that influence the risk of a first primary. The advantages for statistical power are premised on the assumption that relative risks in survivors of a first primary cancer are similar to relative risks in the unaffected population. We explore these assumptions theoretically and empirically using published data from breast cancer studies involving bilateral breast cancer.

Methods—We conducted a literature review to identify case-control studies of variants in 4 genes known to affect breast cancer risk: CHEK2*1100delC; multiple variants in BRCA1 and BRCA2; and FGFR2 rs2981582. Summary odds ratios were obtained for each of three study designs: a conventional case-control design, the design comparing bilateral cases with unilateral controls, and the design comparing bilateral cases with population controls.

Results—The data show strong patterns of steadily increasing prevalence of risk factors from healthy controls to primary cases to bilateral cases, as expected. Relative risks in survivors of unilateral breast cancer are either the same as in the general population, or modestly attenuated.

Conclusions—Patients with double primary malignancies are a very important underused resource for cancer epidemiologic investigation. Such patients are especially useful for broad genome-wide discovery studies, and for studies of rare, strong risk factors such as high penetrance genes.

Epidemiologic enquiry to uncover new risk factors for cancer continues at a fast pace. In recent years the focus has been on genetic risk factors. A few high penetrance genes have been identified, but current thinking is that the risk in individuals is likely to be polygenic in nature, with no way to predict the prevalence of these unknown variants, or the strength of their influences.¹ Current array technology permits the simultaneous study of large numbers of single nucleotide polymorphisms (SNPs), but these focus on relatively common variants.^{2,3} Future sequencing technology is likely to redirect attention towards rarer variants.

In searching the genome for genes and variants that affect risk, it is necessary to use available resources of subjects as efficiently as possible. An idea that has been advocated by a few commentators is to make greater use of patients with multiple primary

malignancies.⁴⁻⁶ The fundamental idea is based on the premise that patients with two or more independent cancers of the same anatomic site represent an exceptionally high-risk group. As such, they must possess considerably increased prevalence of all factors that increase the risk of the cancer under investigation. The use of multiple primaries is especially attractive for studying rare risk factors. It has been shown that, under an assumption that the relative risk is unrelated to the background risk, studies that compare patients with double malignancies to patients with single malignancies can be vastly more powerful for studying rare risk factors than conventional case-control studies.⁴ The increase in statistical power is due simply to the greatly increased relative frequency of the risk variants in the study participants. The power advantages of this design are eliminated when the risk variant is common.

A related idea has been to take advantage of patients with double malignancies by creating case groups that represent an extreme phenotype, and to compare these directly with population controls.^{5,7-9} This “comparison of extremes” approach affects study power by increasing the detectable “signal.” Specifically, if, as before, the relative risk is assumed to be independent of background risk, the odds ratio from the “extremes” design will be the square of the odds ratio for a conventional design. That is, if the odds ratio comparing patients with double primaries to those with single primaries is the same as the odds ratio comparing patients with single primaries to population controls, then the odds ratio from a comparison of double primaries with population controls will be the square of this baseline odds ratio. If true, this increase in signal has the potential to increase the efficiency/power of case-control studies for detecting common variants, such as low penetrance SNPs on a SNP array. In a recent study, Fletcher et al.¹⁰ tested this idea empirically by assembling data from studies of the role of the CHEK2*1100delC variant in causing breast cancer. This large pooled study of 1828 cases of contralateral breast cancer and 7030 population controls produced an odds ratio estimate of 6.4, a result that is consistent with the square of the odds ratio estimate of 2.3 obtained from meta-analyses of conventional studies of the CHEK2*1100delC variant.¹¹

The purpose of the present study is to provide a broader evaluation of the validity of these novel designs. We review the breast cancer literature to uncover reported studies of other known or suspected breast cancer susceptibility genes that have utilized patients with contralateral breast cancer as cases (BRCA1, BRCA2, and FGFR2). Breast cancer appears to be the only malignancy that has been comprehensively analyzed in both its single (unilateral) and multiple (bilateral) forms. Bilateral breast cancer occurs frequently, and as a result is an especially attractive target for this study design. The results are examined to see whether the observed prevalences of the variants among these cases are consistent with the hypothesis that the odds ratio of bilateral breast cancer is the square of the odds ratio of unilateral breast cancer. We also examine the mathematical consequences of this assumption, to determine the potential magnitude of the gains in statistical power of studies that utilize second primaries, contrasting the design implications for common low penetrance variants with rare high penetrance variants. The power advantages of studying second primaries are considered in the light of other theoretical and practical strengths and limitations of these designs.

METHODS

Alternative Design Options and Relative Statistical Power

The study of second primaries as cases has the potential to reduce the required sample size in comparison to a conventional case-control study. This theory was initially presented to justify a study design in which cases comprise incident occurrences of a second primary malignancy.⁴ The ideal control group for such a study would be survivors of a first primary

cancer of the same type. The key assumption of this strategy is that the relative risk due to the risk factor is unrelated to the baseline risk in the population under study, in which case the relative risk would be the same in cancer survivors (a group with a high absolute risk) as in the general population at risk of a first primary. We test this assumption empirically.

We consider three design options and the sample sizes required for each design to deliver equivalent statistical power. Let n_{fc} represent the total sample size for a conventional case-control study, i.e. a study in which first primary cases are compared with population controls. Let n_{sf} represent the sample size for a study of second primary cancers (the “second cancers” design, with first primary cancers as controls) that will deliver equivalent power to the conventional study with sample size n_{fc} . Thus this design is more powerful if n_{sf} is smaller than n_{fc} . Finally, let n_{sc} represent the corresponding sample size for a study in which the cases are second primaries and the controls are unaffected population controls, which we term the “enriched” design. Let the risk factor under investigation have a relative risk in the population denoted by ψ , and let its prevalence be p . Further, for generality we assume that the corresponding relative risk contrasting first and second primaries is ϕ . That is the second-cancers design involves an attenuated relative risk if $\phi < \psi$. Then the sample sizes in the second-cancers design and the enriched design required to deliver the same power as a sample size of n_{fc} for the conventional design are given by

$$n_{sf} = n_{fc} \left[\frac{\log(\psi)}{\log(\phi)} \right]^2 \frac{\phi(1-p+p\psi)^2 + (1-p+p\psi\phi)^2}{\phi(1-p+p\psi)^2 + \psi\phi}$$

$$n_{sc} = n_{fc} \left[\frac{\log(\psi)}{\log(\psi) + \log(\phi)} \right]^2 \frac{\psi\phi + (1-p+p\psi\phi)^2}{\psi\phi + \phi(1-p+p\psi)^2}$$

Further details of the derivation of these formulas are provided in the Appendix. The formulas allow us to specify any sample size (n_{fc}) that will provide an arbitrarily large power for the conventional study, and to use this to compute the required sample sizes for the other design options that will possess equivalent power.

Data Collection

Studies of mechanisms of breast cancer predisposition have found several genetic variants reproducibly associated with risk.^{12,13} These can be broadly categorized as high-risk gene defects (BRCA1, BRCA2), deleterious gene mutations with moderate breast cancer predisposition (CHEK2*1100delC), and low-penetrance gene polymorphisms (for example, FGFR2 rs2981582). We also considered a polymorphic haplotype suggested as a possible risk factor (ATM composite allele ins38(-8) and 5557A), but after reviewing the very limited literature^{14–16} we elected not to present the results of the ATM haplotype because there is no persuasive evidence that this variant affects breast cancer risk. For BRCA1 and BRCA2 we reviewed studies that involved mutations with proven deleterious effect on the gene’s function. We examined the common BRCA1 variant 5382insC separately, as a few studies examined this mutation in isolation. Each of these categories of predisposing alleles has been studied with various levels of scrutiny in bilateral breast cancer cases, unilateral breast cancer cases, and non-affected controls.

We identified case-control studies of these selected genes and the risk of breast cancer published before 1 March 2009, through computer-based searches of PubMed. Using FGFR2 as an example, we performed two consecutive searches as follows: (1) (“breast cancer” OR “breast neoplasms” OR “breast carcinoma”) AND (FGFR2 OR rs2981582) NOT review [pt]; (2) (“bilateral breast cancer” OR “multiple breast cancer” OR “multiple primary cancer”) AND (FGFR2 OR rs2981582) NOT review [pt]. This was repeated for the other genes and variants by replacing “FGFR2 OR rs2981582” with “BRCA1 OR

BRCA2”, and “CHEK2”. Studies were included if they involved women with unilateral breast cancer or bilateral breast cancer as cases and unilateral breast cancer or breast cancer-free as controls. We further restricted the studies to those for which there was no evidence that the participants were selected on the basis of a family history of breast cancer. We also included case-series studies (without controls) or studies consisting solely of unaffected individuals, as these contribute to our estimates of the population prevalences of the genes in one of the three comparison groups (i.e. population controls, unilateral breast cancer cases, bilateral breast cancer cases) even though these studies do not contribute to our summary odds ratio estimates. We extracted from each study information on study design, geographic location, ethnicity, age and numbers of cases and controls (eTable 1, <http://links.lww.com>).

Statistical Analysis

The mutation frequencies were compared for heterogeneity across studies using the Fisher-Freeman-Halton Test (StatXact©).¹⁷ We applied conventional meta-analytic techniques to groups of studies employing the same study design (i.e. conventional, second-cancers or enriched).¹⁸ This involved checks for between-study heterogeneity¹⁹ and publication bias,²⁰ and calculation of summary odds ratios (ORs) and 95% confidence intervals (CIs). We used the statistical software STATA (version 10.0; STATA Corp, College Station, TX) and StatXact (Cytel Corporation, Cambridge, MA).

RESULTS

Our search identified 21 studies of CHEK2*1100delC, 20 studies of various BRCA1 truncating mutations (7 in Ashkenazi populations, 13 in other populations), 12 reports involving solely the BRCA1 5382insC variant (5 in Ashkenazi populations, 7 in other populations), 17 studies of various BRCA2 truncating mutations (7 in Ashkenazi populations, 10 in other populations), and 3 studies of FGFR2 rs2981582. Detailed results from each of these studies are provided in eTable 2 (<http://links.lww.com>) along with the results of the various statistical analyses. These include heterogeneity tests of the prevalence estimates of the variants in each of the three key groups (population controls, unilateral breast cancer, and bilateral breast cancer), summary odds ratios from meta-analyses of the studies for each of the different study types (conventional case-control design; second-cancers design; enriched design), and tests of heterogeneity and publication bias for the various summary odds ratio estimates.

Summary results are provided in Table 1 by gene. The overall relative frequencies show the anticipated pattern of increasing prevalence from population controls to unilateral cases to bilateral cases. For example, for CHEK2*1100delC the estimated frequencies of the variant rise from an average of 0.5% in the populations studied to 1.5% in unilateral cases to 2.8% in bilateral cases. These prevalence estimates vary considerably from study to study (note heterogeneity *P*-values), reflecting possible differences in frequencies among ethnically distinct populations. However, for genes with large effects, such as BRCA1 and BRCA2, these variations are minor relative to the differences between the case groupings.

In analyzing the odds ratios, these variations across studies justify stratification by individual study, as is conventional in any meta-analysis. However, the heterogeneity tests for the meta-analyses of odds ratios are all non-significant. This provides some reassurance about the validity of the summary odds ratio estimates (right 3 columns of Table 1), although the relatively small numbers of component studies provides limited power for these heterogeneity tests. Note that these analyses have combined studies of Ashkenazi and non-Ashkenazi for both BRCA1 and BRCA2, because the odds ratios appear to be consistent. These summary odds ratios confirm our thesis that associations in conventional case-control

investigations (unilateral cases vs control column) will also be detected in studies that compare bilateral cases with unilateral cases.

The odds ratio for BRCA1 appears to be attenuated in the bilateral vs unilateral design, although for the other three genes the estimates with those 2 designs are very similar. The results also support the corollary to this thesis, that the enriched design (bilateral vs control column) provides a stronger signal. For CHEK2*1100delC and FGFR2 the estimates from this design are close to the square of the odds ratio estimates from the corresponding conventional studies. For BRCA1 and BRCA2 the odds ratio estimates from this design actually exceed the square, although there are few studies and the confidence intervals are wide. All of these comparisons are limited by the relatively few studies available for meta-analysis, especially for BRCA2 and FGFR2, and the wide confidence intervals throughout.

The power implications of these findings are shown in Table 2. This table provides the relative effective sample sizes of the three study designs required to achieve a given level of statistical power, using a conventional study with 1000 subjects as the reference. The parameters represent various risk-factor prevalences similar to the ones under investigation, including a common variant (prevalence 0.3), an uncommon variant (prevalence 0.01) and a very uncommon variant (prevalence 0.001), and for modest (odds ratio of 1.5) and high strength (odds ratio of 5.0) associations. For example, under the first set of parameter assumptions, a study comparing double primaries with population controls would require only 240 participants to achieve equivalent power. (The absolute power differs from row to row in the table, so it is not meaningful to compare results between rows within a column.)

In the top half of the table it is assumed that the relative risk of disease for a given risk factor for the ratio of second primaries to first primaries is not attenuated compared with the ratio of first primaries to ordinary controls. The bottom half of the table repeats the results assuming that the relative risk of disease for a given risk factor is reduced 25% for the ratio of second primaries to first primaries. If there is no attenuation of relative risk with second primaries (top half of table), the second-cancers design is more efficient than a conventional design for rarer and stronger risk factors, i.e. fewer participants are required to achieve equivalent study power). This advantage is reduced if the signal is attenuated (bottom half of table). Across the board, the enriched design has greater power than the conventional design.

DISCUSSION

In using second primary cancers for epidemiologic studies of cancer risk, the fundamental premise is that any risk factor that increases the risk of the cancer under study in people previously unaffected with the disease will also increase the risk of a second primary among cancer survivors. Our results support this hypothesis for genetic risk factors associated with breast cancer. This hypothesis assumes there is nothing unique or etiologically distinct about the occurrence of double malignancies (i.e. that they represent two independent occurrences of the disease), and two occurrences will typically occur in people at high risk. Given this presumption, we have explored the further hypothesis that the degree of elevation of risk for any risk factor is similar (on a multiplicative scale) in the setting of cancer survivors to what is observed in the general population. Our results suggest that on the relative scale (ie using relative risks as opposed to, say, risk differences) the risk elevation among cancer survivors is typically similar, although with possible attenuation in some cases. Of the 4 genes investigated, only one (BRCA1) had a summary relative risk in the second cancers design substantially smaller than the summary estimate from the case-control studies. In contrast, the 3 “enriched” studies of BRCA1 produced a summary odds ratio higher than expected (62). A modest attenuation of the relative risk is consistent with an investigation of known risk factors for melanoma in a similar setting.²¹ The possibility of such attenuation has

implications for statistical power. Our calculations show that studies involving second primaries have major advantages in terms of statistical power if the relative risk can be considered to be constant. Attenuation of the relative risk can alter this balance of power, depending on the magnitude of the attenuation. However, studies that compare second primaries with population controls have power advantages even with attenuation, and across the spectrum of risk factor characteristics.

Our results provide only an imprecise investigation of these phenomena. Many of the individual studies in the literature are vague with respect to criteria for case and control selection. Just as with conventional case-control studies, the other two designs would ideally be constructed using population-based sampling of both cases and controls. Also, these designs have practical merit only for selected cancer sites, i.e. those for which second primaries in the same organ type are common and clearly distinguishable from metastatic lesions. This includes cancers of the breast, lung, colon-rectum and skin, but excludes rare cancers and those for which much of the primary site is typically removed by surgery, e.g. prostate cancer. Also, multiple primaries are common for head and neck cancers, but reliable discrimination between multiple primaries and superficial metastases is complicated.

In what circumstances might the relative risk of any risk factor be different in cancer survivors than in the general population? One possibility (as just discussed) is diagnostic error whereby metastases are misdiagnosed as second primaries. There is a substantial literature of studies evaluating this issue, but the consensus for breast cancer is that most contralateral occurrences are indeed independent occurrences of the disease.²²⁻²⁴ Recent studies support this conclusion for melanoma, but suggest that mis-diagnoses may be common for multiple primary lung cancers.²⁵⁻²⁷ Another possibility is interaction with treatment. Common treatments for primary breast cancer include agents such as tamoxifen that reduce the incidence of the disease by 50%. If the sub-types of tumors that are prevented by treatment are associated with a genetic risk factor then the impact of this risk factor overall will be different in second primaries compared with first primaries. These influences could affect the relative risk, but they are unlikely to affect the detectability of any risk factor. Another possibility is simply that the relative effect of individual risk factors diminishes as the background risk increases. Indeed studies showing an approximately constant risk in BRCA1/2 carriers by age would seem to support this thesis, in that the “background” risk increases markedly with age.⁶ A final possibility is that the variant may be associated with case survival. If so, the odds ratio could be either attenuated or enhanced, depending on the direction of this association.

Many investigators studying genetic risk factors elect to “enrich” the case selection by restricting attention to cases with a family history of the disease. This approach is similarly designed to increase power by genetically enriching the case base. These studies are rarely population-based, and it is more difficult to make a precise estimate of the extent to which the power is likely to be enhanced. Interestingly, Antoniou and Easton²⁸ have studied this issue using a polygenic model with a normally distributed polygenic component estimated from a large population-based study; they conclude that if one restricts case selection to breast cancer cases with an affected mother and sister, the study will deliver increased power of a similar order of magnitude to using cases with bilateral breast cancer. Of course, a consideration in the use of any “enriched” design (whether using cases with multiple primary or cases with a strong family history of cancer) is the added difficulty in identifying and recruiting these enriched cases. Our statistical power comparisons assume equivalent sample sizes but do not address the relative ease by which these can be obtained in practice.

The conventional case-control study has a long history in cancer epidemiology. The gold standard is the population-based design, whereby incident cases from a defined population

are compared with controls randomly selected from the same population. However, this ideal is increasingly challenged by the difficulty in enrolling population controls with a high response rate.²⁹ While it is suitable for relatively common risk factors, the design is problematic for important but rare risk factors (ie those that confer a high relative risk). An enriched design based on second primaries offers an attractive alternative in that it provides substantial power advantages across a broad spectrum of risk-factor prevalences and relative risks. However, like the conventional approach, it requires population controls with the attendant difficulties. An enriched design may be especially attractive for genome-wide association studies of candidate SNPs due to its power advantages for a broad spectrum of SNP prevalences. The second-cancers design, by its case-only nature, promises higher participation rates, especially when biologic samples are required (as for genetic analysis). This design has more statistical power than the conventional design for rare risk factors, although its advantages can be muted if there is risk attenuation. Its power is competitive with that of the enriched design for very rare, strong risk factors. This is likely to be an increasingly important area as knowledge of strong genetic risk factors emerges. Major genes such as BRCA1 and BRCA2 possess hundreds of individual variants, many extremely rare, and consequently individual studies require a high level of power to distinguish harmful rare variants from the harmless ones.³⁰

In summary, our study provides strong empirical evidence that studies of second primary cancers are capable of detecting cancer risk factors, in many circumstances with greatly improved power. This underused resource could be employed to facilitate the on-going search for cancer risk factors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Sources of financial support: Supported by an International Cancer Fellowship from the International Union Against Cancer (Application No ICR/08/140), by the National Cancer Institute (Awards CA131010 and CA124504), and by the Russian Foundation for Basic Research (grant numbers 09-04-90402; 08-04-00369-1; 07-04-00122) and the Government of Moscow (grant number 15/09).

We thank Peter Devilee and Petra Huijts (Leiden University Medical Center) for access to original data from their study.

References

1. Pharoah PD, Antoniou A, Bobrow M, et al. Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet.* 2002; 31:33–36. [PubMed: 11984562]
2. Easton DF, Pooley KA, Dunning AM. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature.* 2007; 447:1087–1093. [PubMed: 17529967]
3. Hunter DJ, Kraft P, Jacobs KB. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet.* 2007; 39:870–874. [PubMed: 17529973]
4. 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]
5. Imyanitov EN, Cornelisse CJ, Devilee P. Searching for susceptibility alleles: emphasis on bilateral breast cancer. *Int J Cancer.* 2007; 121:921–923. [PubMed: 17444500]
6. 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]
7. Kuligina ES, Togo AV, Suspitsin EN, et al. CYP17 polymorphism in the groups of distinct breast cancer susceptibility: comparison of patients with the bilateral disease vs. monolateral breast cancer

- patients vs. middle-aged female controls vs. elderly tumor-free women. *Cancer Lett.* 2000; 156(1): 45–50. [PubMed: 10840158]
8. Suspitsin EN, Buslov KG, Grigoriev MY, et al. Evidence against involvement of p53 polymorphism in breast cancer predisposition. *Int J Cancer.* 2003; 103(3):431–3. [PubMed: 12471629]
 9. Buslov KG, Iyevleva AG, Chekmariova EV, et al. NBS1 657del5 mutation may contribute only to a limited fraction of breast cancer cases in Russia. *Int J Cancer.* 114(4):585–9. [PubMed: 15578693]
 10. Fletcher O, Johnson N, Dos Santos Silva I, et al. Family history, genetic testing, and clinical risk prediction: pooled analysis of CHEK2 1100delC in 1,828 bilateral breast cancers and 7,030 controls. *Cancer Epidemiol Biomarkers Prev.* 2009; 18:230–234. [PubMed: 19124502]
 11. CHEK2 Breast Cancer Case-Control Consortium. CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from 10 studies. *Am J Hum Genet.* 2004; 74:1175–1182. [PubMed: 15122511]
 12. Antoniou AC, Pharoah PD, McMullan G, et al. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *Br J Cancer.* 2002; 86:76–83. [PubMed: 11857015]
 13. Garcia-Closas M, Hall P, Nevanlinna H, et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet.* 2008; 4:e1000054. [PubMed: 18437204]
 14. Angèle S, Romestaing P, Moullan N, Vuillaume M, Chapot B, Friesen M, Jongmans W, Cox DG, Pisani P, Gerard JP, Hall J. ATM haplotypes and cellular response to DNA damage: association with breast cancer risk and clinical radiosensitivity. *Cancer Res.* 2003; 63:8717–25. [PubMed: 14695186]
 15. Langholz B, Bernstein JL, Bernstein L, Olsen JH, Børresen-Dale AL, Rosenstein BS, Gatti RA. The WECARE Study Collaborative Group, Concannon P. On the proposed association of the ATM variants 5557G>A and IVS38-8T>C and bilateral breast cancer. *Int J Cancer.* 2006 Aug 1; 119(3):724–5. [PubMed: 16496408]
 16. Tommiska J, Jansen L, Kilpivaara O, Edvardsen H, Kristensen V, Tamminen A, Aittomäki K, Blomqvist C, Børresen-Dale AL, Nevanlinna H. ATM variants and cancer risk in breast cancer patients from Southern Finland. *BMC Cancer.* 2006 Aug 16; 6:209. [PubMed: 16914028]
 17. Freeman GH, Halton JH. Note on an exact treatment of contingency, goodness of fit and other problems of significance. *Biometrika.* 1951; 38:141–149. [PubMed: 14848119]
 18. Sutton, AJ.; Abrahms, KR.; Jones, DR., et al. *Methods for Meta-Analysis in Medical Research.* Wiley; New York: 2000.
 19. Dersimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials.* 1986; 7:177–188. [PubMed: 3802833]
 20. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics.* 1994; 50:1088–1101. [PubMed: 7786990]
 21. Begg CB, Hummer AJ, Mujumdar U, et al. A design for cancer case-control studies using only incident cases: experience with the GEM study of melanoma. *Int J Epidemiol.* 2006; 35:756–764. [PubMed: 16556646]
 22. Saad RS, Denning KL, Finkelstein SD, et al. Diagnostic prognostic utility of molecular markers in synchronous bilateral breast carcinoma. *Mod Pathol.* 2008; 21:1200–1207. [PubMed: 18469799]
 23. Tse GM, Kung FY, Chan AB, et al. Clonal analysis of bilateral mammary carcinomas by clinical evaluation and partial allelotyping. *Am J Clin Pathol.* 2003; 120:168–174. [PubMed: 12931545]
 24. Imyaninov EN, Suspitsin EN, Grigoriev MY, et al. Concordance of allelic imbalance profiles in synchronous and metachronous bilateral breast carcinomas. *Int J Cancer.* 2002; 100:557–564. [PubMed: 12124805]
 25. Orlow I, Tommasi DV, Bloom B, et al. Evaluation of the clonal origin of multiple primary melanomas using molecular profiling. *J Invest Dermatol.* 2009; 129:1972–82. [PubMed: 19282844]
 26. Girard N, Ostrovskaya I, Lau C, et al. Genomic and mutational profiling to assess clonal relationships between multiple non-small cell lung cancers. *Clin Cancer Res.* 2009; 15:5184–90. [PubMed: 19671847]

27. Wang X, Wang M, MacLennan GT, et al. Evidence for common clonal origin of multifocal lung tumors. *J Natl Cancer Inst.* 2009; 101:560–570. [PubMed: 19351924]
28. Antoniou AC, Easton DF. Polygenic inheritance of breast cancer: Implications for design of association studies. *Genet Epidemiol.* 2003; 25:190–202. [PubMed: 14557987]
29. Olson SH, Voigt LF, Begg CB, et al. Reporting participation in case-control studies. *Epidemiology.* 2002; 13:59–65. [PubMed: 11805587]
30. Capanu M, Orlov I, Berwick M, et al. The use of hierarchical models for estimating relative risks of individual genetic variants: an application to a study of melanoma. *Stat Med.* 2008; 27:1973–1992. [PubMed: 18335566]

Appendix

The goal is to develop concise formulas to characterize the relative power (efficiency) of the three candidate designs, denoted conventional, second primaries and enriched. We use the following notation.

Subjects

- Conventional healthy controls (subscript c): prevalence of risk factor = p
- First primary cases (subscript f): prevalence of risk factor = q
- Second primary cases (subscript s): prevalence of risk factor = r

Designs

We assume that the “true” relative risk of the risk factor in the general population at risk is ψ , but that this could be different (e.g. attenuated) in the population of cancer survivors. Setting this attenuated relative risk to be ϕ , the detectable “signals” for the various designs are as follows:

- Conventional Design: comparison of f versus c: signal = $\log(\psi)$
- Second Primaries Design: comparison of s versus f: signal = $\log(\phi)$
- Enriched Design: comparison of s versus c: signal = $\log(\psi) + \log(\phi)$

Variances

Each design ultimately involves calculation of the odds ratio linking the risk factor with case-control status. The variances of these odds ratios are in effect the variances of the estimates of the signals above. These variances are functions of the prevalences of the risk factor in the relevant comparison groups, and also of the underlying relative risks, ψ and ϕ . Recognizing that $q = p\psi/(1-p + p\psi)$ and $r = p\psi\phi/(1-p + p\psi\phi)$, it follows that the three variances are as defined below:-

Conventional Design:

$$v_{fc} = n_{fc}^{-1} \left(\frac{1}{p} + \frac{1}{1-p} + \frac{1}{q} + \frac{1}{1-q} \right) = n_{fc}^{-1} \frac{\psi^2 + \psi(1-p + p\psi)^2}{p\psi^2(1-p)}$$

where n_{fc} is the number of subjects per group (cases, controls), assumed to be of equivalent size for convenience;

Second Primaries Design:

$$v_{sf} = n_{sf}^{-1} \left(\frac{1}{q} + \frac{1}{1-q} + \frac{1}{r} + \frac{1}{1-r} \right) = n_{sf}^{-1} \frac{\varphi(1-p+p\psi)^2 + (1-p+p\psi\varphi)^2}{p\psi\varphi(1-p)};$$

Enriched Design:

$$v_{sc} = n_{sc}^{-1} \left(\frac{1}{p} + \frac{1}{1-p} + \frac{1}{r} + \frac{1}{1-r} \right) = n_{sc}^{-1} \frac{\psi\varphi + (1-p+p\psi\varphi)^2}{p\psi\varphi(1-p)}.$$

Relative Efficiency

The power of any of these designs is dependent on the ratio of the signal to the standard error signal of the estimate of the signal. This is because the power takes the form

$\Phi(-z_\alpha + \frac{\text{signal}}{v^{1/2}})$, where v is the relevant variance from above. Thus to obtain the relative efficiency of any two designs we need to determine the relative sample sizes required to

achieve equivalent $\frac{\text{signal}}{v^{1/2}}$ targets. These are obtained simply by equating the corresponding signal to noise ratios. Consequently the sample sizes required to achieve equivalent power are given by

$$n_{sf} = n_{fc} \left[\frac{\log(\psi)}{\log(\varphi)} \right]^2 \frac{\varphi(1-p+p\psi)^2 + (1-p+p\psi\varphi)^2}{\varphi(1-p+p\psi)^2 + \psi\varphi}$$

$$n_{sc} = n_{fc} \left[\frac{\log(\psi)}{\log(\psi) + \log(\varphi)} \right]^2 \frac{\psi\varphi + (1-p+p\psi\varphi)^2}{\psi\varphi + \varphi(1-p+p\psi)^2}.$$

Table 1

Summary Frequencies of Variants and Relative Risks

Variant	Summary Relative Frequencies; % (heterogeneity p-values)		Summary Odds Ratios (95%CI) ^a [Test for heterogeneity P-value; publication bias ^b ; no. studies]	
	Controls	Breast cancer	Unilateral breast cancer vs Control	Bilateral breast cancer vs Control
		Unilateral	Unilateral vs bilateral breast cancer	Bilateral breast cancer vs Control
Chek2*1100delC	0.5 (<0.01)	1.5 (<0.01)	2.6 (2.0–3.3) [0.95; 0.87; 17]	6.0 (3.5–10.2) [0.51; 0.85; 5]
BRCA1 mixed				
Ashkenazi	1.1 (0.48)	7.7 (<0.01)		
Not Ashkenazi	0.3 (0.08)	2.0 (<0.01)		
5382insC				
Ashkenazi	0.2 (0.20)	1.9 (0.03)		
Not Ashkenazi	0.4 (0.21)	2.1 (<0.01)	6.3 (4.7–8.3) [0.40; 0.57; 6]	62 (12–317) [0.47; --; 3]
BRCA2 mixed				
Ashkenazi	1.2 (0.05)	3.3 (0.97)		
Not Ashkenazi	0.0 (–)	1.0 (0.27)	3.3 (2.5–5.2) [0.74; --; 2]	30 (1.6–562) [–; --; 1]
FGFR2				
TT vs CC	15.5 (<0.01)	19.6 (<0.01)	1.5 (1.3–1.8) [0.80; --; 3]	2.2 (1.5–3.4) [0.54; --; 2]
TC vs CC	46.1 (<0.01)	48.7 (0.02)	1.0 (0.7–1.4) [0.42; --; 2]	1.3 (1.0–1.9) [0.68; --; 2]

^a Estimates and confidence intervals calculated using a “fixed effects” approach.

^b Test for publication bias performed only for analyses with ≥ 5 studies.

Table 2

Effective Sample Sizes by various parameters for 3 study designs

Prevalence (p)	Parameters		Effective Sample Sizes; no.		
	Relative Risk (ψ)	Attenuated (ϕ)	Conventional	Second Cancers	Enriched
0.3	1.5	1.5	1000	920	240
0.3	5.0	5.0	1000	1860	470
0.01	1.5	1.5	1000	670	220
0.01	5.0	5.0	1000	230	220
0.001	1.5	1.5	1000	670	220
0.001	5.0	5.0	1000	200	220
0.3	1.5	1.375	1000	1486	307
0.3	5.0	4.0	1000	2689	754
0.01	1.5	1.375	1000	1132	280
0.01	5.0	4.0	1000	318	254
0.001	1.5	1.375	1000	1121	280
0.001	5.0	4.0	1000	284	253