Design and Analysis Issues in a Population-Based, Case–Control–Family Study of the Genetic Epidemiology of Breast Cancer and the Co-operative Family Registry for Breast Cancer Studies (CFRBCS)

John L. Hopper, Georgia Chenevix-Trench, Damien J. Jolley, Gillian S. Dite, Mark A. Jenkins, Deon J. Venter, Margaret R. E. McCredie, Graham G. Giles

Background: **Historically, studies of the "genetic epidemiology" of cancer have used nonsystematically sampled kindreds with numerous cases of cancer across multiple generations. From the epidemiologic viewpoint, it is difficult to extrapolate findings to the population because of the *ad hoc* ascertainment of these atypical, ill-defined families. Since 1992, we have been conducting a population-based, case–control–family study of breast cancer.** ***Methods:*** **Families are identified through a single, population-sampled proband, who is either affected or unaffected, making adjustment for ascertainment straightforward. Administered questionnaires and blood samples are sought from cases, controls, and specified sets of relatives. From 1996 through 1999, a further 1200 case families have been recruited as part of the Co-operative Family Registry for Breast Cancer Studies (CFRBCS). Issues relevant to the study design and analysis are discussed.** ***Results:*** **Epidemiologic and genetic findings published to date are summarized. In particular, this population-based study has shown that the so-called "high-risk" families containing multiple cases of breast cancer are *not* typical of families in the general population in which BRCA1 or BRCA2 mutations are segregating. Most "hereditary" cancers are "sporadic."** ***Conclusion:*** **The collection of DNA, as well as data on disease status and risk factors, from population-sampled sets of relatives provides a powerful resource for addressing genetic and environmental determinants of cancer. A population-based multicenter, multidisciplinary enterprise, such as has been developed by the CFRBCS, may become a model for future research in cancer epidemiology, allowing genetic and environmental risk factors to be put into a proper population perspective. [Monogr Natl Cancer Inst 1999;26:95–100]**

In 1982, Morton *(1)* defined the new discipline of "genetic epidemiology" by modifying the definition of "epidemiology" given by Last *(2)*. It was to be "a science that deals with the etiology, distribution, and control of disease in groups of relatives and with inherited causes of disease in populations." In this context, "inherited" was meant to include biologic and nonbiologic inheritance *(1)*. That is, in genetic epidemiology the basic unit of study is the family (as distinct from the individual), and both genetic and nongenetic causes of familial aggregation are to be considered. Clearly, the method of sampling families (i.e., ascertainment) is a key issue.

Over the past few decades, however, studies purporting to be of the genetic epidemiology of cancer have almost exclusively used families in which there are a number of cases of one or more types of cancer, spread across multiple generations. This strategy has been important, as it has led to the identification of "familial cancer syndromes" and of some of the genes causing cancer in a proportion of these atypical kindreds. From an epidemiologic viewpoint, it is difficult to extrapolate findings from these families—such as estimates of disease risk in mutation carriers—to any defined set of people, let alone the general population. This is because of the lack of a clear or consistent definition of sampled families and because of the *ad hoc* (i.e., nonrandom) ascertainment, violating the basic principles of statistical inference. Whereas these genetic studies have been important, it is difficult to classify them as part of "genetic epidemiology," where "epidemiology" is a noun, and "genetic" is an adjective meant to also include consideration of nongenetic factors *(1)*.

To be able to make valid conclusions about the roles of genetic factors in regard to breast cancer in the population and to put the roles of genetic and nongenetic factors into their proper perspective, we have been conducting a population-based study of the genetic epidemiology of breast cancer *(3,4)*. In this paper, we will describe the study and discuss some design and analysis issues based on our experiences.

Australian Breast Cancer Family Study (ABCFS)

Families are identified through a single, population-sampled proband, who is either affected or unaffected. This sampling makes adjustment for ascertainment straightforward, simply by conditioning on the proband's age and disease status. Case probands are women recently diagnosed with a first primary, invasive breast cancer reported to the Victorian or New South Wales Cancer Registry and living in Melbourne or Sydney, respectively, at the time of their diagnosis. Cancer registration is compulsory under legislation and is considered to be complete. Con-

Affiliation of authors: J. L. Hopper, D. J. Jolley, G. S. Dite, M. A. Jenkins, Centre for Genetic Epidemiology, The University of Melbourne, Carlton, Australia; G. Chenevix-Trench, Queensland Institute of Medical Research, and Department of Pathology, University of Queensland, Herston, Australia; D. J. Venter, Department of Pathology and Research, Peter MacCallum Cancer Institute, Melbourne, Australia; M. R. E. McCredie, New South Wales Cancer Council, Woolloomooloo, Australia, Department of Preventive and Social Medicine, University of Otago, Dunedin, New Zealand; G. G. Giles, Cancer Epidemiology Centre, The Anti-Cancer Council of Victoria, Carlton, Australia.

Correspondence to: John L. Hopper, Ph.D., The University of Melbourne, Centre for Genetic Epidemiology, 200 Berkeley St., Carlton, Victoria 3053, Australia (e-mail: j.hopper@gpph.unimelb.edu.au).

trol probands are women without breast cancer, selected at random from the electoral rolls for the relevant cities (adult registration for voting is compulsory in Australia). Questionnaires are used to measure epidemiologic risk factors in case probands, control probands, and relatives of *both* types of probands. We call this a "case–control–family" design *(3)*. Blood samples are collected from case probands and control probands and from specified sets of relatives depending on the family cancer history. In this design, strict rules define which relatives are to be studied (*see below*).

Why Did We Do the Study?

At the time the ABCFS commenced (initial grants were written in the late 1980s, and pilot work and recruitment began in 1992), the expression "BRCA1" had yet to be coined. Our primary aims at the time were to collect the best possible epidemiologically valid data to study the effect of family history on risk of breast cancer and to obtain a resource of DNA from population-based families for future research once genes associated with breast cancer susceptibility were identified. Previously, the typical epidemiologic approach had been to rely solely on unverified self-reports collected as part of conventional case–control and cohort studies in which family history was only one of a large number of putative risk factors being addressed.

We also wished to apply the new "regressive logistic" models for analysis of family data *(5,6)* to test empirically the extent to which familial aggregation in risk factors might account for familial aggregation of cancer. [A paper on the theoretical aspects of this issue was published in 1992 *(7)*]. These new statistical methods, based on logistic regression and fitted by maximum likelihood, also allow determination of the extent to which the *residual* familial aggregation can be attributed to an *unmeasured* genetic effect (commonly referred to as a "major gene effect"). The residual familial aggregation is that remaining after adjusting for risk factors *measured* on individual family members, with or without adjustment for the effect of parental disease status on an individual's risk, and for disease dependencies within sibships *(5)*. The risk factors could be measured genetic variants, measured environmental or lifestyle factors, or a combination of both. For example, if mutation status for the currently identified breast cancer susceptibility genes, BRCA1 and BRCA2, is known for case probands and at least some of their relatives, this design and approach to statistical analysis will allow us to address questions such as how much familial aggregation is explained by BRCA1 and BRCA2 mutations; what evidence is there for other "major breast cancer genes"; and what is the most likely penetrance and prevalence of mutations in these, as yet, unknown genes?

To optimize information on genetic factors, we started by sampling women with early onset breast cancer, i.e., breast cancer diagnosed before the age of 40 years. Case–control and cohort studies have consistently indicated that the multiplicative increase in risk to close female relatives of an affected woman is greater for early onset disease, and the earlier the onset of disease in the affected relative. Furthermore, segregation analyses have suggested that the proportion of breast cancer attributable to a major gene is greater the earlier the onset of disease. In recent years, we have been able to extend the sampling to case probands with later age at diagnosis (*see below*).

Why Did We Not Use a Two-Stage Design, Based on Oversampling Women With Breast Cancer Who Self-Report Having a Family History?

Although oversampling affected individuals with a family history of the disease will be more efficient for studying familial factors *(8,9)*, in practice this strategy will become problematic if the decision to sample is based on incorrect information. Previous experience in Australia with a population-based family study of malignant melanoma found that, for a high percentage of respondents, their knowledge of malignant melanoma in relatives could not be relied on; the false-positive reporting of family history by affected case probands was 40% *(10)*. Although we did not expect the problem of relying on reports of disease in relatives to be of the same magnitude for breast cancer, it served as a warning, particularly as we wanted to extend our pedigrees to more than first-degree relatives of probands. In practice, studying all cases irrespective of family history has made conduct and analysis easier, but it may have led to inefficiencies in data collection for some analyses. However, given that the study was being developed as a resource to address a number of hypotheses, some of which would involve genes and genetic markers not yet discovered, the extra cost and effort may in the long term prove to be worthwhile.

Why Study Population-Based Control Families?

Epidemiology is built around the comparison of affected and unaffected subjects. Therefore, we considered that, as we were studying relatives of affecteds (case families), we should also study relatives of unaffecteds (control families). Furthermore, and according to the principles of modern epidemiology, case families and control families should be studied using as far as practical identical methods. This approach was motivated by our principal concerns: to measure the population cancer burden associated with genes; to study concurrently the roles and possible interactions of both genetic and environmental risk factors on cancer susceptibility; and to identify genetic and environmental factors and, in particular, modifiable lifestyle factors that ameliorate or exacerbate susceptibility to cancer in women at high genetic risk. Such information is necessary to determine the relative costs and benefits of different genetic screening programs and to help plan optimal strategies.

Whom to Interview and From Whom to Take a Blood Sample?

A number of design issues were addressed during the conduct of the study. Major questions revolved around how wide to extend the core family (i.e., whether or not to include aunts, uncles, grandparents, nephews, and nieces) and who to take blood from and who to interview.

In pilot work, we studied the case proband and her sisters, brothers, parents, and grandparents. It was soon realized that, given the young age at diagnosis, only a small proportion had an affected female relative (<15%). Therefore, although it meant a substantial increase in the number of interviews per family, it was decided to extend the sampling frame to include aunts.

Table 1 shows, for each type of relative, the proportions for whom a personal interview or a proxy interview was conducted and the proportion deceased, for the relatives of the 248 women with breast cancer diagnosed before age 40 years, studied between 1992 and 1995 and living in Melbourne. For most of this period, all sisters of cases were asked to provide a blood sample,

Table 1. Proportions of relatives of cases diagnosed before the age of 40 years, by vital status and interview status, for the Melbourne component of the Australian Breast Cancer Family Study, 1992–1995

Relative	Alive, %		Deceased, proxy, %	No data, %
	Interviewed	Proxy		
Sisters	76	22	1	1
Brothers	60	35	3	2
Mothers	65	21	14	—
Fathers	41	26	33	—
Aunts	39	36	19	4
Grandmothers	3	9	85	3
Grandfathers	1	3	92	4

and this was achieved for 60% of identified sisters. For each case, there were on average 1.4 sisters, 1.4 brothers, and 2.8 aunts.

Proxy interviews were used for persons who were deceased or who refused to participate so as to collect information on important risk factors, such as parity, that could reasonably be asked of close relatives. Male relatives were interviewed because they may be at an increased risk of breast or other cancers (at the start of the study there was little epidemiologic information to support such a conjecture, but the increased risk of breast and prostate cancers now apparently evident in male carriers of mutations in BRCA2, and BRCA1 and BRCA2, respectively, supports this decision). The male questionnaire is the same as for females, excluding questions relevant only to females.

It can be seen from Table 1 that the proportions interviewed were highest for mothers and sisters, lower for males and for the older generations of relatives, being lowest for grandparents for whom the great majority were deceased. These interview proportions are likely to be lower in the relatives of older case probands.

Of the total of 644 eligible case probands from Melbourne and Sydney studied between 1992 and 1995 *(4)*, 467 were interviewed (72.5%). Attrition was a result of death (1.7%), refusal (surgeon 8.4%; patient 11.8%), nonresponse (surgeon 0.6%; patient 1.4%), or having moved residence and not being located (3.6%). Of the 632 eligible control probands, refusals (25.8%) and nonresponse (9.8%) resulted in 408 being interviewed (64.4%). Blood samples were available from 393 case probands (84.2% of participating and 61.0% of those eligible) and 294 control probands (72.1% of participating and 46.5% of those eligible) [*see (11)*].

The final protocol for whom to interview and from whom to obtain a blood sample was as follows: In the first instance, blood is sought from both the case and control probands as well as from the mother, father, and any sisters of the case probands. (Blood is sought only from relatives of control probands if they have two or more affected relatives.) Interviews are sought from all adult first-degree relatives and from aunts and grandparents of either sex on both sides of the family. In case families, should a first- or second-degree relative be reported to have been affected with breast or ovarian cancer, blood and an interview are sought from the affected relative and all first-degree relatives from whom blood has not already been sought. Should any of these newly approached relatives be affected, blood and interviews are sought from their yet-to-be approached first-degree relatives, and so on. This sampling scheme is a valid version of the Cannings–Thompson Sequential Ascertainment Scheme in which adjustment for ascertainment is achieved by conditioning on the disease status of the proband *(12)*.

Co-operative Family Registry for Breast Cancer Studies (CFRBCS)

Since 1996, we have studied a further 1200 case families as part of the National Institutes of Health-funded CFRBCS, an international resource for studies of the genetic epidemiology of breast cancer. This resource consists of two other centers focusing on population-based studies similar to the one described above (one in San Francisco, CA, and the other in Ontario, Canada) and three U.S. centers recruiting families with extensive histories of breast cancer through family cancer clinics. Core questionnaires and uniform protocols for the collection and storage of blood and tissue have been developed. Pathology review is under way, so that issues concerning the comparability of diagnoses of cancer subtypes within and between centers and countries can be addressed. Pathology has become a more important issue following our population-based confirmation of distinct histologic phenotypes *(13)* and molecular pathogenic pathways *(14)* in breast carcinomas occurring in women with a germline mutation in BRCA1 or BRCA2.

Some Further Design Issues

A number of design issues have arisen out of the first few years' experience in establishing and conducting the CFRBCS, a multicenter project in genetic epidemiology. One important issue revolves around the need, at least for some exposures, to ask probands about their experience up until a pre-designated time. For case probands, we chose "1 year before diagnosis" as their reference age. For control probands, we have asked about their experiences up until the time of interview. In an unmatched case–control study of probands, we made the reference age for control probands their age at interview and adjusted for reference age in the analysis *(4)*.

The reference age issue is also critical when case probands are compared with unaffected sisters in a matched analysis. This is because it is possible that the sibling controls may have changed their behavior(s) following the diagnosis of the proband, or there may be secular trends in the exposure (as was historically the case for oral contraceptive use). Therefore, it will be necessary to truncate information on the exposure of sister(s) to the calendar date of the case proband's diagnosis. If, following the principles of risk-set sampling, only unaffected sisters who have attained the reference age of the case proband are eligible, they would need to have done so before the calendar time of diagnosis. Analyses can then be conducted with the use of truncated or nontruncated exposures to assess to what extent the effect of learning of one's sister's diagnosis might influence the results. These and a number of similarly important issues in using unaffected sibling or family controls are discussed in a companion paper in this monograph *(15)*.

The issue of how to word questions must also recognize that, although the study has a case–control component, it is also a family study and has the potential of becoming a cohort study of families in the long term. Therefore, questionnaire design must also take into account the likelihood of follow-up questionnaires being administered at regular intervals and updating exposures in both affected and unaffected individuals as well as the vital and disease status of all relatives. Items that need to be considered include who to follow-up and the method(s) of follow-up

(e.g., active follow-up using mailed questionnaires, telephone interviews, etc., or passive follow-up relying on identification of cancer and deaths with the use of population registries, etc.). Note that, when family history data are updated, this must be done systematically to protect from any bias that might arise from more attention being given to those families in which one is told that new cases have occurred.

Statistical Estimation of Cumulative Risk

Estimation of cumulative risk by analyzing cancer histories in a defined set of relatives of mutation carriers in which, as in our study, the mutation status of all relatives is not necessarily known poses an interesting statistical challenge. We have used two approaches: one based on repeated sampling and Kaplan–Meier curves *(16)* and the other based on maximum likelihood estimation assuming a functional form for the cumulative risk *(17)*. These methods were used to derive the penetrance estimate from analysis of time to diagnosis of breast cancer in known and potential mutation-carrying relatives and were shown to give almost identical estimates and confidence intervals *(16)*.

The likelihood approach can also be applied to estimate the risk of disease associated with other genes, even those for which the increased risk may be modest. This approach is made possible because of our defined population-based ascertainment scheme for which correct adjustment can be made, the use of population incidence rates derived by the registries from which case probands have been sampled, and the quality of the data on disease status that derives from our seeking to interview living relatives and attempting to verify all reports of cancers in the family *(4)*. Furthermore, the mode of inheritance can be specified as dominant, recessive, or codominant. The estimation of risk in terms of the hazard ratio permits interpretation in the usual case–control manner, allowing direct estimation of the attributable risk associated with genetic variants within each age category. It is not necessary to measure genetic variants for all—or even any—relatives, and there is no bias from missing genetic information, provided there is no nonpaternity or nonmaternity, because the genetic status of unmeasured relatives is inferred probabilistically, conditional on the measured genotype of the proband and that of any relative. The likelihood approach can, at least in theory, be extended to include other causes of familial aggregation and to incorporate risk factors measured on relatives, as discussed above in reference to regressive logistic models.

Results

Some Findings to Date: Epidemiology

Between 1992 and 1995, the ABCFS studied 467 women with breast cancer and 408 control probands. A standard case–control analysis *(4)* revealed that the main risk factor was having a first-degree relative with breast cancer; *see also* Table 2 in which the categories have been expanded and one case proband since found to be ineligible has been deleted. The great majority of affected first-degree relatives were mothers, and this risk factor was associated with an odds ratio of about 3. Having a sister affected was associated with a greater odds ratio, but, because of the paucity of affected sisters of control probands, the estimate had a wide confidence interval. The increased risk associated with at least one affected second-degree relative was about 1.3 but not nominally significant, although only effects greater than twofold could be excluded with reasonable statistical power.

Table 2. Estimates of familial risks of breast cancer in women before the age of 40 years, based on reported history of breast cancer in female relatives from the Australian Breast Cancer Family Study, 1992–1995

Affected relative(s)	% (No.) affected		Relative risk* (95% confidence interval)
	Case subjects	Control subjects	
Mother	10.7 (50)	4.9 (20)	2.9 (1.6–5.2)
Sister	3.2 (15)	0.3 (1)	16.4 (2.1–128)
First-degree	12.5 (58)	5.2 (21)	3.4 (1.9–5.9)
Maternal aunt	5.6 (26)	6.9 (28)	0.9 (0.5–1.6)
Paternal aunt	7.5 (35)	4.2 (17)	1.8 (1.0–3.4)
Any aunt	12.2 (57)	11.0 (45)	1.2 (0.8–1.9)
Maternal grandmother	6.4 (30)	5.6 (23)	1.2 (0.7–2.2)
Paternal grandmother	4.7 (22)	2.9 (12)	1.6 (0.7–3.4)
Any grandmother	10.9 (51)	8.3 (34)	1.4 (0.9–2.3)
Second-degree	21.7 (101)	18.6 (76)	1.3 (0.9–1.8)
Second-degree only	17.8 (83)	17.9 (73)	1.0 (0.7–1.4)
First- or second-degree	30.3 (141)	23.0 (94)	1.6 (1.2–2.2)

*Adjusted for state, reference age, country of birth, education, marital status, age at menarche, number of live births, breast lump—benign, and height [*see (4)*].

As well as family history, other factors were found to be predictive of case–control status, but their effects were clearer when modeled together in a multiple logistic regression, fitting linear effects across ordered categories of exposure rather than different risk estimates for each category *(4)*.

The odds ratios and relative risk estimates for all those risk factors considered to make up the parsimonious model are shown in Table 3. Having ever had a child, on its own, appeared to be weakly predictive and of marginal statistical significance *(4)*. Once the number of live births was taken into account, however, it was seen that the first birth was associated with a 65% increase in risk but that each subsequent birth reduced the risk by 23% [*see* Fig. 1 in *(4)*]. Current use of oral contraceptives was associated with an increase in risk of 36%, but this finding was of marginal significance.

Height was associated with an increased risk of about 3% per cm [*see* Fig. 2A in *(4)*]. There was limited evidence of a decreased risk with weight or body mass index (BMI), as had previously been reported in the literature. Fitting height and weight (or height and BMI) together, however, enhanced the height effect and suggested a protective effect of weight limited to women in the highest quintile [*see* Fig. 2B in *(4)*].

Table 3. Estimates of log odds ratios and relative risk for the risk factors included in the parsimonious model for the Australian Breast Cancer Family Study, 1992–1995

Risk factor	Log odds ratio* (standard error)	Relative risk* (95% confidence interval)
Ever had live birth	0.50 (0.22)	1.65 (1.06–2.56)
Number of children—1	−0.26 (0.10)	0.77 (0.63–0.94)
Current oral contraceptive use	0.31 (0.17)	1.36 (0.97–1.92)
Height, cm	0.027 (0.011)	1.03 (1.005–1.050)
Affected first-degree relative	1.05 (0.27)	2.86 (1.67–4.90)

*Adjusted for state, reference age, country of birth, education, and marital status [*see (4)*].

We were aided in these analyses and interpretations by use of floating-point standard errors *(18)* that present indications of uncertainty of risk estimates for each category while taking into account adjustments for other risk factors; *see* Figs. 2A and 2B in *(4)*. We calculated approximate standard errors with the use of a simple formula based on an heuristic argument (details available from the first author on request).

We were motivated in our "linear modeling of effects" in part by parsimony and in part by our future intentions to examine whether the effects of these epidemiologic risk factors depend on, or are "modified by," genetic factors. The power to detect such interactions or modifications of risk will be optimized if the nongenetic effect can be well summarized by just one parameter.

Some Findings to Date: Genetic

Some genetic analyses have now been undertaken in the DNA extracted from blood samples collected by the ABCFS between 1992 and 1995. For example, it had been suggested that the codon 325 polymorphism in the estrogen receptor (ER) gene was predictive of breast cancer in women with a family history of breast cancer *(19)*. That observation was based on a series of affected subjects only, selected from a clinic and for whom family cancer history was based on self-report and not verified. Our much larger study *(11)*, using population-based sampling and trying to validate family history going back two generations, found no supporting evidence despite having sufficient statistical power to detect effects one-half the magnitude observed in the hypothesis-generating study. We also found no evidence for an effect of the androgen receptor exon 1 CAG repeat length on risk of breast cancer *(20)*. Finally, using new methods to size minisatellite alleles, we found no evidence to support an effect of the rare alleles at the HRAS1 locus on the risk of early onset breast cancer, despite having ample statistical power to detect an effect of the size predicted by a meta-analysis of previous studies of breast cancer with the use of a less sensitive technique to classify alleles *(21)*.

We have also been conducting extensive mutation testing in these population-based samples *(16,22)*. We estimated the average risk of breast cancer in women who carry a protein-truncating mutation in BRCA1 or BRCA2, the supposedly "high-risk genes for familial breast cancer," with the use of new statistical approaches *(16,17)*. Our estimate, based on studying the relatives of those cases with breast cancer diagnosed before the age of 40 years who were found by us to carry a germline protein-truncating mutation, was about one half that derived from analyses of data from atypical multiple-case families *(17)*.

The striking finding of our study is that the so-called "high-risk" families that contain multiple cases of breast cancer are *not* typical of the families in the general population in which BRCA1 or BRCA2 mutations are segregating. More often than not, the mutation-carrying case diagnosed before the age of 40 years does not have a family history of breast cancer, even going back two generations. It will be of interest to see if this finding holds for case subjects with later age at diagnosis.

Fig. 1 illustrates schematically the intermixing of "familial," "sporadic," and "hereditary" cases of breast cancer in the population, based on our mutation testing. "Familial" cases are those with an affected first- or second-degree relative at the time of diagnosis. "Sporadic" cases are those with no such known family history. As shown in Table 4, the sporadic cases comprise about 70% of all women under the age of 40 years at diagnosis *(4)*, and about one in 20 of these cases carry a BRCA1 or BRCA2 mutation, i.e., are "hereditary." In total, there are more hereditary cases among the sporadic group than among the familial group! In about 28% of all early-onset cases, only one relative is affected. About one in 16 of these familial cases

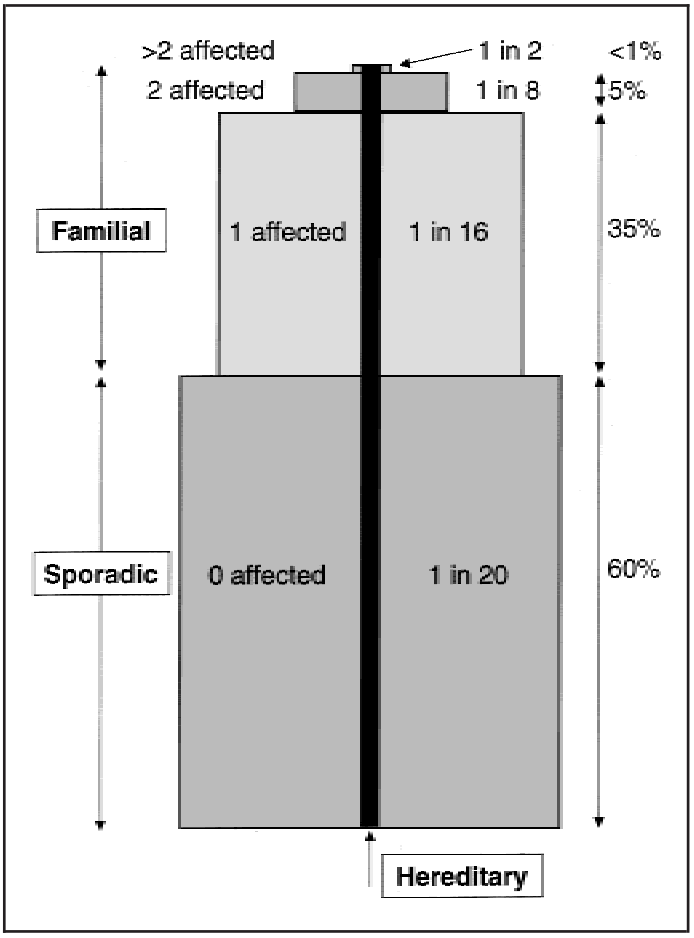

Fig. 1. Women with breast cancer diagnosed before the age of 40 years, categorized according to number of affected first- or second-degree female relatives (0, 1, 2, or >2). The relative areas of the **boxes** indicate the proportion of all affected cases in each category of family history. "Familial" cases have one or more affected first- or second-degree relatives and make up about 30% of all cases *(4)*. "Sporadic" cases have no family history. "Hereditary" cases have a germline mutation in either BRCA1 or BRCA2 and can occur in any category. Although they are more likely to occur if the case is familial, in absolute numbers, most hereditary cases are sporadic. The proportion of hereditary cases in each category of family history, in terms of 1 in *x*, is shown in the **right-hand half of each box.** The proportion of all hereditary cases for each category of family history, in terms of a percentage, is shown on the **right-hand side of the figure.**

Table 4. Approximate breakdown of mutation status in BRCA1 or BRCA2 for women with breast cancer diagnosed before age 40 years, by family history, based on the Australian Breast Cancer Family Study, 1992–1995

No. of affected first- or second-degree female relatives	Proportion of all cases, %	Proportion who carry a mutation in BRCA1/2	Proportion of all mutation carriers, %
0	70	~1 in 20	60
1	28	~1 in 16	35
2	2	~1 in 8	5
>2*	<1	~1 in 2	<1

*Based on the Breast Cancer Linkage Consortium [*see (17)*].

carries a mutation, the more so if the other affected family member is a first-degree relative (one in 10) than a second-degree relative (one in 20). If two relatives are affected, then about one in eight carries a mutation, but the latter group constitutes only 2% of all cases. Cases with three or more affected relatives make up less than 1% of all cases, even those of early onset, and less than 1% of all mutation-carrying cases. [In Fig. 1, we have presumed that about one half of these carry a mutation in either BRCA1 or BRCA2, based on the findings of the Breast Cancer Linkage Consortium; see *(17).*] That is, so-called "high-risk families" constitute less than 1% of all BRCA1 or BRCA2 mutation carriers, whereas 60% of mutation carriers are in families with *no* family history.

Fig. 1 resembles the shape of a three-story tower with a broad base representing the sporadic cases and a small top section representing those with the strongest family history. It will be of interest to find out what the shape of this tower looks like for case patients with a later age at diagnosis as well as for other cancers and other genes. Does the tower become fatter or thinner, do the familial components get bigger or smaller, and does the heritable core get wider or narrower?

Our findings have implications for genetic testing and counseling, raising the possibility that risk prediction in mutation carriers may need to vary according to previous family history of breast or other cancers. Clearly, systematic prospective studies of multiple-case mutation-carrying families are needed to assess if they really deserve to be classified as "high risk." It also raises the interesting prospects that cancer risk may vary considerably between mutations, even between those that are protein-truncating, or that genetic and environmental risk factors may exist that modify risk in mutation carriers. Extensive mutation testing in large population-based family studies such as ours, with information on both genetic and questionnaire-derived risk factors, will be important in clarifying these issues.

Conclusion

Population-based multicenter, multidisciplinary enterprises, such as those being collected by the CFRBCS, may become a model for the future research in cancer epidemiology and cancer genetics. Cancer epidemiology will benefit by moving from considering the individual to the family as the unit of interest and, therefore, allowing assessment of the role of genetic factors, especially if blood or tissue samples, or both, are available. Cancer genetics will benefit immeasurably by the fact that findings can be immediately put into a population and clinical perspective.

References

(1) Morton NE. Outline of genetic epidemiology. Basel (Switzerland): Karger; 1982.

(2) Last J. A dictionary of epidemiology. Oxford (UK): Oxford University Press; 1983.

(3) Hopper JL, Giles GG, McCredie MR, Boyle P. Background, rationale and protocol for a case–control–family study of breast cancer. Breast 1994;3: 79–86.

(4) McCredie MR, Dite GS, Giles GG, Hopper JL. Breast cancer in Australian women under the age of 40. Cancer Causes Controls 1998;9:189–98.

(5) Bonney GE. Regressive logistic models for familial disease and other binary traits. Biometrics 1986;42:611–25.

(6) Hopper JL, Jenkins MA, Macaskill GT, Giles GG. Regressive logistic modelling of familial aggregation for smoking in a population-based sample of nuclear families. J Epidemiol Biostats 1997;2;45–52.

(7) Hopper JL, Carlin JB. Familial aggregation of a disease consequent upon correlation between relatives in a risk factor measured on a continuous scale. Am J Epidemiol 1992;136:1138–47.

(8) Whittemore AS, Halpern J. Multi-stage sampling in genetic epidemilogy. Stat Med 1997;16:153–67.

(9) Siegmund KD, Whittemore AS, Thomas DC. Multistage sampling for disease family registries. Monogr Natl Cancer Inst 1999;26:43–8.

(10) Aitken JF, Youl P, Green A, MacLennan R, Martin NG. Accuracy of case-reported family history of melanoma in Queensland, Australia. Melanoma Res 1996;6:313–7.

(11) Southey MC, Batten LE, McCredie MR, Giles GG, Dite G, Hopper JL, et al. Estrogen receptor polymorphism at codon 325 and risk of breast cancer in women before age forty. J Natl Cancer Inst 1998;90:532–6.

(12) Thompson EA. Pedigree analysis in human genetics. Baltimore (MD): The Johns Hopkins University Press; 1986.

(13) Armes JE, Egan AJ, Southey MC, Dite GS, McCredie MR, Giles GG, et al. The histologic phenotypes of breast carcinoma occurring before age 40 years in women with and without BRCA1 or BRCA2 germline mutations: a population-based study. Cancer 1998;83:2335–45.

(14) Armes JE, Trute L, White D, Southey MC, Hammet F, Tesoriero A. et al. Distinct molecular pathogeneses of early-onset breast cancers in BRCA1 and BRCA2 mutation carriers: a population-based study. Cancer Res 1999; 59:2011–7.

(15) Gauderman WJ, Witte JS, Thomas DC. Family-based association studies. Monogr Natl Cancer Inst 1999;26:31–7.

(16) Hopper JL, Southey MC, Dite GS, Jolley DJ, Giles GG, McCredie MR et al. for the Australian Breast Cancer Family Study. Population-based estimate of the average age-specific cumulative risk of breast cancer for a defined set of protein-truncating mutations in BRCA1 and BRCA2. Cancer Epidemiol Biomark Prevent 1999;8:741–47.

(17) Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, et al. and the Breast Cancer Linkage Consortium. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. Am J Hum Genet 1998;62:676–89.

(18) Easton DF, Peto J, Babiker AG. Floating absolute risk: an alternative to relative risk in survival and case–control analysis avoiding an arbitrary reference. Stat Med 1991;10:1025–35.

(19) Roodi N, Bailey LR, Kao WY, Verrier CS, Yee CJ, Dupont WD, et al. Estrogen receptor gene analysis in estrogen receptor-positive and receptor-negative primary breast cancer. J Natl Cancer Inst 1995;87:446–51.

(20) Spurdle AB, Dite GS, Chen X, Mayne CJ, Southey MC, Batten LE, et al. Androgen receptor exon 1 CAG repeat length is not a risk factor for breast cancer in women under the age of 40. J Natl Cancer Inst 1999;91:961–6.

(21) Firgaira FA, Seshadri R, McEvoy CR, Dite GS, Giles GG, McCredie MR, et al. HRAS1 rare minisatellite alleles and breast cancer in Australian women under age forty years. J Natl Cancer Inst 1999;24:2107–11.

(22) Southey MC, Tesoriero AA, Andersen CR, Jennings KM, Brown SM, Dite GS, et al. BRCA1 mutations and other sequence variants in a population-based sample of Australian women with breast cancer. Br J Cancer 1999; 79:34–9.