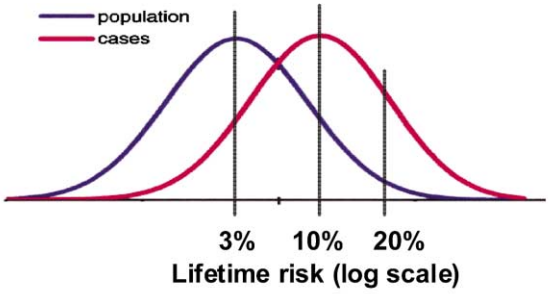
Breast cancer susceptibility—A new look at an old model

A polygenic model in which many individually weak genes combine multiplicately to cause a 50-fold range of risk in the population explains several puzzling aspects of familial breast cancer epidemiology, including the very high risk in some families and the failure to identify important new genes since the discovery of *BRCA1* and *BRCA2*.

Germline mutations in *BRCA1* or *BRCA2* account for most large multiple-case breast cancer families but only a few percent of unselected cases. It has been recognized since *BRCA2* was cloned that these and a few other known genes account for less than a quarter of the observed excess risk in relatives of breast cancer patients (Ponder, 2001), yet no important new breast cancer gene has been discovered in the intervening six years despite extensive linkage and candidate gene studies. Systematic sequencing in patients and controls has identified a few candidate polymorphisms that may confer risk ratios of the order of 1.5 or 2 (Dunning et al., 1999), but most are of marginal statistical significance, and even in combination they cannot make a substantial contribution to the overall familial effect. Ponder and his colleagues, writing in *Nature Genetics* (Pharoah et al., 2002), discuss the possibility that there may be few, if any, major breast cancer genes still to be discovered. They propose a polygenic basis for the wide observed variation in familial risk whereby the individually small effects of many genes combine multiplicatively to produce a log-normal distribution of lifetime risk in the general population, as shown by the blue curve in Figure 1. Lifetime risk varies more than 50-fold, from less than 1 in 200 for women below the 3rd percentile to more than 20% above the 97th percentile. The genetic distribution of risk among breast cancer patients also spans a 50-fold range, but their average risk is 4-fold greater than in the general population (red curve). If each gene conferred a 1.5-fold risk, the number of such genes that would be needed to account for the doubling of risk observed in patients' relatives ranges from several hundred if their individual frequencies were 0.01 to a few dozen with a population frequency of 0.1 (Ponder, 2001). The model was developed by segregation analysis in a population-based series of families and in a collection of multiple case families in which breast cancer patients had been screened for *BRCA1* and *BRCA2* mutations (Antoniou et al., 2002). The assumption that the hypothetical "polygenes" multiply the effects of all other genes including *BRCA1* and *BRCA2* accounts for the higher penetrance seen in *BRCA* carriers from multiple-case families than in those identified by screening unselected cases. The analysis also supports the suggestion that most breast cancers arise in a susceptible minority of women (Peto and Mack, 2000). Half of all breast cancers are predicted to occur in the most susceptible 12% of the population.



Lifetime risk	Prevalence in the general population	Prevalence among breast cancer cases
<3%	0.50	0.12
3-10%	0.38	0.38
10-20%	0.09	0.25
>20%	0.03	0.25

Figure 1. Polygenic model for breast cancer risk

Distribution of lifetime breast cancer risk in the general population and in women who will develop breast cancer (figure based on Pharoah et al., 2002).

This classical polygenic model is uniquely defined by the relative risk in patients' relatives. The predicted relative risk of 4 in patients (and in their identical twins) is the square of the observed relative risk in their sisters, which is about 2. Its predictions, particularly the distribution of multiple cases in families, should therefore provide a critical test of the model's plausibility. In practice, however, several quite different genetic models provide a statistically satisfactory fit to such family data. The original CASH model involved a rare highly penetrant dominant gene (Claus et al., 1991), and models that combine the *BRCA* genes with a single less-penetrant recessive or dominant gene appear to fit virtually as well as the polygenic model (Antoniou et al., 2002). (The marginally higher risk in patients' sisters than in their daughters probably reflects shared nongenetic factors.) Such genes might also interact with *BRCA1* and *BRCA2* to account for their higher penetrance in multiple-case families. This resurrection of the classical polygenic model is timely and important because the hunt for more penetrant genes has proved fruitless and it is time to rethink our strategy, not because the details are likely to be biologically correct. The statement that "the validity of our results depends on the validity of the segregation analysis, a detailed critique of which is beyond the scope of this paper" gives the misleading impression that such model-fitting can provide reliable evidence on underlying mechanisms. The suggestion that because the relative risk in patients' relatives is also about 2 for many other common cancers their distribution of risk is likely to exhibit similar variation (Pharoah et al., 2002) is also questionable. The relative risk in people with two affected relatives is less than 3 for breast cancer, but for most other cancers it is between 10 and 30 (Dong and Hemminki, 2001), suggesting that a smaller proportion of cases occur in highly susceptible individuals. But their central thesis, that important advances in breast cancer genetics will depend on the discovery of large numbers of genes with weak effects, is likely to be correct, and gene hunters will ignore it at their peril.

These authors had already reported

the fit of this polygenic model to their breast cancer families (Antoniou et al., 2002), and this article is concerned mainly with the implications of a genetic test that might in the future identify individuals at very high risk as well as those at such low risk that they would not want to be screened at all. The eventual identification and functional characterization of a large number of such genes would certainly have a major impact on many aspects of breast cancer prevention and treatment, but the more immediate question is how they can be found. Coincidentally, the same issue of *Nature Genetics* includes a report that a single truncating mutation in *CHEK2* found in 1.1% of the general population and only 1.4% of unselected breast cancers is carried by 5% of patients from multiple-case families unlinked to *BRCA1* or *BRCA2* (*CHEK2*-Breast Cancer Consortium, 2002). The contribution of mutant *CHEK2* to overall breast cancer incidence is trivial, but the demonstration that rare low-penetrance genes can be found by simply comparing their prevalence in patients with affected relatives against controls is an important methodological advance. (*CHEK2* fails to exhibit the hypothesized interaction with *BRCA1* or *BRCA2*, probably because it lies in the same functional pathway.) Such genes may be almost impossible to detect by traditional approaches, including linkage in families or prevalence in unselected cases. Most susceptibility genes will not carry such a common mutant haplotype, and the search for further genes will probably be restricted to candidates for the next year or two. When rapid genome-wide sequencing becomes available, however, any genetic variant with a frequency of 1% or more that confers a risk of 1.5–2 should be detectable by this approach.

The recent overview of risks in patients' relatives based on 58,209 women with breast cancer (Collaborative Group on Hormonal Factors in Breast Cancer, 2001) provides the most precise published estimates of familial risks. The incidence in a patient's relatives is further increased almost 2-fold if another relative is also affected, confirming that there is considerable heterogeneity in underlying incidence among women who develop breast cancer. The polygenic model provides a plausible explanation of this variation in lifetime risk. But it does not explain why the risk in patients' relatives at ages older than the index case's age at diagnosis is independent of the index case's age (Collaborative Group on Hormonal Factors in Breast Cancer, 2001), or why contralateral rates are roughly constant at all ages. The high rate at young ages in their relatives shows that young patients are genetically different from older patients, yet their relatives and those of older patients eventually suffer virtually identical age-specific rates. Different combinations of genes may determine a woman's eventual level of risk and the age at which that risk is reached (Peto and Mack, 2000), but such speculative hypotheses cannot be tested until the genes that underlie these surprising patterns have been cloned and characterized. The polygenic model is an extremely useful conceptual simplification, but the interest will lie in the complexity.

Julian Peto

Institute of Cancer Research
Sutton, Surrey SM2 5NG
and London School of Hygiene and
Tropical Medicine
London WC1E 7HT
United Kingdom
E-mail: j.peto@icr.ac.uk

Selected reading

Antoniou, A.C., Pharoah, P.D.P., McMullan, G., Day, N.E., Stratton, M.R., Peto, J., Ponder, B.J., and Easton, D.F. (2002). Br. J. Cancer *86*, 76–83.

CHEK2-Breast Cancer Consortium (2002). Nat. Genet. *31*, 55–59.

Claus, E.B., Risch, N., and Thompson, W.D. (1991). Am. J. Hum. Genet. *48*, 232–242.

Collaborative Group on Hormonal Factors in Breast Cancer (2001). Lancet *358*, 1389–1399.

Dong, C., and Hemminki, K. (2001). Int. J. Cancer *92*, 144–150.

Dunning, A.M., Healey, C.S., Pharoah, P.D.P., Teare, M.D., Ponder, B.A.J., and Easton, D.F. (1999). Cancer Epidemiol. Biomarkers Prev. *8*, 843–854.

Peto, J., and Mack, T.M. (2000). Nat. Genet. *26*, 411–414.

Pharoah, P.D.P., Antoniou, A., Bobrow, M., Zimmern, R.L., Easton, D.F., and Ponder, B.A.J. (2002). Nat. Genet. *31*, 33–36.

Ponder, B.A.J. (2001). Nature *411*, 336–341.