Table 1

Prevalence of the Three BRCA1 and BRCA2 Founder Mutations in Ashkenazi Jewish Women Who Have Never Had a Diagnosis of Breast or Ovarian Cancer, as a Function of Family History of Breast Cancer and Age at Testing (Hartge et al. [1999], table 3)

	Proportion (%) of Individuals with BRCA1/BRCA2 Founder Mutations When				
	No. of First-Degree Relatives with Breast Cancer Is			Early Diagnosis in Affected Relatives Is	
Age at Testing (years)	0	1	⩾2	Not Present	Present
<40	9/566 (1.6)	10/123 (8.1)	0/1 (0)	5/62 (8.1)	5/62 (8.1)
40–49	14/888 (1.6)	9/217 (4.1)	0/7 (0)	3/143 (2.1)	6/81 (7.4)
50–59	8/636 (1.3)	4/163 (2.5)	2/12 (17)	4/119 (3.4)	2/56 (3.6)
⩾60	4/615 (.7)	1/163 (.6)	1/28 (3.6)	1/124 (.8)	1/67 (1.5)
Total	35/2,705 (1.3)	24/666 (3.6)	3/48 (6.3)	13/448 (2.9)	14/266 (5.3)

Am. J. Hum. Genet. 65:1771, 1999

Modeling the Probability That Ashkenazi Jewish Women Carry a Founder Mutation in BRCA1 or BRCA2

To the Editor:

The Washington study (see Struewing et al. 1997) currently provides the closest to population-based data on the prevalence of the three common BRCA1 (MIM 113705) and BRCA2 (MIM 600185) founder mutations in Ashkenazi Jewish women, so the most use must be made of it. In this respect, Hartge et al. (1999) should be congratulated for presenting raw data in their recent publication in the *Journal*.

Hartge et al. (1999) use the classification-and-regression tree (CART) approach to identify the "most important" predictors of mutation status. They conclude that a previous diagnosis of breast or ovarian cancer was the most important discriminator. For women with such a cancer history (i.e., affected individuals), the next most important predictor was apparently age at diagnosis, after which "family history discriminated relatively little" (p. 965). In contrast, for men and women without a personal history of breast or ovarian cancer (i.e., unaffected individuals), family history "best discriminated" between carriers and noncarriers. In their table 2, Hartge et al. present the most important subcategories of risk, derived by CART, along with the observed frequency in each cell. The number of carriers in some cells is small, however, making some inferences imprecise.

The CART method for ranking predictors in terms of "importance" is influenced by the distribution of predictive factors, not just by the size of the effect itself. It is of interest, therefore, to estimate the *strengths* of putative predictive factors with regard to the probability of being a carrier, both for affected and for unaffected individuals. One way is by the use of multiple logistic regression, in which the effects are estimated on the log-odds scale. Not only does this statistical approach give explicit estimates of the magnitude of the effects of predictors, along with measures of the precision of these estimates (such as standard errors [SEs] and confidence intervals [CIs]), it also allows one to summarize the data in terms of as few parameters as is possible for extraction of the most information. Such a "parsimonious" model uses all the available information to produce a general picture of the strength and nature of the major predictive factors. (Hartge et al. [1999] do not report the findings of logistic-regression analyses, apparently because they found that the estimated frequencies were "substantially higher" than the observed frequencies in groups with observed prevalence >20%. As can be seen from their tables 3 and 4, however, this only applied to 34 cases at age <40 years at diagnosis, 20 of age 40–49 years at diagnosis, and 10 of age 50–59 years at diagnosis, i.e., <2% of the 3,742 women in the total data set.)

Tables 1 and 2 present the observed prevalence of mutation carriers, extracted from tables 3 and 4 of Hartge et al. (1999), both for women with and for women without a diagnosis of *breast* cancer, by age at diagnosis or age at testing. It also categorizes women according to their family history of *breast* or *ovarian*

Table 2

Prevalence of the Three BRCA1 and BRCA2 Founder Mutations in Ashkenazi Jewish Women Who Have Had a Diagnosis of Breast Cancer, as a Function of Family History of Breast Cancer (in a First-Degree Relative) and Age at Diagnosis (Hartge et al. [1999], table 4)

	Proportion (%) of Individuals with BRCA1/BRCA2 Founder Mutations When				
	No. of First-Degree Relatives with Breast Cancer Is			Early Diagnosis in Affected Relatives Is	
Age at diagnosis (years)	0	1	⩾2	Not Present	Present
<40	6/27 (22)	3/7 (43)	...	1/3 (33)	2/4 (50)
40–49	6/77 (8)	3/25 (12)	2/7 (29)	2/23 (9)	3/9 (33)
50–59	2/59 (3)	3/21 (14)	1/2 (50)	2/15 (13)	2/8 (25)
⩾60	1/47 (2)	0/22 (0)	0/3 (0)	0/15 (0)	0/10 (0)
Total	15/210 (7)	9/75 (12)	3/12 (25)	5/56 (9)	7/31 (23)

cancer, defined either by the number of first-degree relatives with breast cancer or by whether any first-degree relative had breast cancer diagnosed at age <50 years (referred to there as "early diagnosis"). Several points are readily seen:

1. The prevalence increases for each affected first-degree relative, for both affected *and* unaffected individuals; for example, as the number of affected relatives increases from 0 to 1 to 2, the prevalence in affected individuals increases from 1 in 14 to 1 in 8 to 1 in 4; in unaffected individuals, it increases from 1 in 80 to 1 in 28 to 1 in 16.

2. In women with a family history, the prevalence is greater if at least one of the affected relatives had an early diagnosis, for both affected and unaffected individuals.

3. There is a substantial difference, in prevalence between affected and unaffected individuals; this applies to each of the family-history and age categories, although it is not necessarily of the same strength in each category.

4. There is evidence for an "age" effect in both affected *and* unaffected individuals; for example, in women with a family history of breast cancer, the prevalence decreases from 43% to 16% to 8% as age *at diagnosis* increases from age <40 to 40–49 to ⩾50 years. In unaffected individuals, it decreases from 8% to 4% to 2% as the age *at testing* increases from age ⩽40 to 40–49 to ⩾50 years.

Multiple logistic-regression analyses were conducted for the data in tables 1 and 2, by use of the software GLIM (Baker and Nelder 1978). Likelihood-ratio tests were used to test between nested models, and the scaled deviance was used as a goodness-of-fit test (McCullagh and Nelder 1983).

First, family history was analyzed in terms of the number of affected first-degree relatives. (Note that there were no women of age <40 years in the data set who both were affected and had more than one affected first-degree relative, so this category has been deleted from the analysis.) In neither affected nor unaffected individuals was there evidence for an interaction between the effects of family history and "age" ($\chi^2_5 = 3.3$, $P = .6$; and $\chi^2_6 = 6.2$, $P = .4$, respectively). Furthermore, the effects of both age and family history were well represented by one parameter each: a linear effect on the log-odds scale per each age category (both $\chi^2_2 < 2$, $P < .4$) and per each affected relative (both $\chi^2_1 < 0.1$, $P < .8$). Within each group, there was no evidence that a logistic model that involved such linear effects for age and number of affected relatives gave an unacceptable fit ($\chi^2_9 = 7.5$, $P = .6$; $\chi^2_8 = 5.3$, $P = .7$, respectively). On the log-odds scale, the effect of age at diagnosis in affected individuals was -0.970 (SE, 0.257) per age category ($P < .001$), and the effect of age at testing in unaffected individuals was -0.434 (0.130) per age category ($P < .001$). The effect per affected relative was 0.826 (0.327) in affected individuals ($P = .01$) and 1.061 (0.222) in unaffected individuals ($P < .001$); that is, this family history effect was no different between affected and unaffected individuals ($P = .6$).

When data from affected and unaffected individuals were combined, the parsimonious model gave an acceptable fit ($\chi^2_{18} = 13.2$), and the linear predictor (with SEs of regression parameters in parentheses) was

$$\begin{aligned} &0.986\,(0.185) \times \text{no. of affected relatives} + \\ &[-4.319\,(0.380) - 0.426\,(0.129) \\ &\quad \times \text{age at testing in unaffecteds}], \text{ or} \\ &[-1.432\,(0.606) - 0.993\,(0.257) \\ &\quad \times \text{age at diagnosis in affecteds}] \,. \end{aligned} \tag{1}$$

(Here age is categorized as 1–4, respectively, for the groups age <40, 40–49, 50–59, and ⩾60 years.) The effect of having one affected relative was 0.983 (0.232), no different from that of having a second, which was 0.956 (0.503); $P > .9$. The average effect per affected relative was equivalent to the odds of being a mutation carrier increasing by $\exp(0.986) = 2.7$-fold (95% CI, 1.9–3.9) for each affected first-degree relative, irrespec-

tive of whether the woman was affected. In affected individuals, it decreased by 0.4-fold (95% CI, 0.2–0.6) per category of age at diagnosis, whereas in unaffected individuals, it decreased by 0.7-fold (95% CI, 0.5–0.8) per category of age at testing.

By a simple reparameterization, it can be shown that the odds were greater in affected than in unaffected individuals with the same number of affected relatives, but the difference depended on age at diagnosis/testing, decreasing from ~10-fold (95% CI, 3.7–28), for women of age <40 years, to 1.9-fold (95% CI, 1.4–2.4), for women of age ⩾60. The CIs indicate that the effect of being affected was clearly significant at all ages.

Second, the same analysis was carried out again, this time with women with a family history categorized as 0, for no relatives affected; 2, for having at least one relative with "early diagnosis"; and 1 otherwise. As before, in neither affected nor unaffected individuals was there evidence for an interaction between the effects of family history of breast cancer and age ($\chi^2_6 = 3.3$, $P = .6$; $\chi^2_6 = 4.0$, $P = .7$, respectively). Furthermore, the effect of family history was well represented by a linear effect on the log-odds scale, per each family-history category (both $\chi^2_2 < 0.4$, $P = .8$), that was no different between affected and unaffected individuals; 0.752 (0.272) versus 0.741 (0.155), $P > .9$.

Combining affected and unaffected individuals, the parsimonious model again gave an acceptable fit ($\chi^2_{19} = 10.7$, $P = .9$), and the linear predictor was

$$\begin{aligned} &0.744\ (0.135) \times \text{family history category} + \\ &[-4.184(0.371) - 0.383(0.128) \\ &\quad \times \text{age at testing in unaffected individuals}], \text{ or} \\ &[-1.199(0.582) - 0.862(0.251) \\ &\quad \times \text{age at diagnosis in affected individuals}]\ ; \end{aligned} \qquad (2)$$

that is, the estimated effects were very similar to those in equation (1). (The effect of family history was 0.775 [0.287] in going from no family history to category 1, no different from the effect of 0.705 [0.341] in going from category 1 to category 2; $P > .8$.)

For both equation (1) and equation (2), there was no evidence for a poor fit, in that the estimated probabilities were not unduly disparate from the observed prevalences, in any category. In particular, for the two categories with the highest estimated probabilities according to equation (1), .39 in both cases, the observed values were .43 and .29 from samples of $n = 7$ each. For equation (2), the three highest estimated probabilities were .52, .34, and .29, similar to the observed prevalences of .50, .33, and .33 from samples of $n = 4$, 3, and 9, respectively.

The estimated probability of being a carrier, derived from the parsimonious models (1) and (2) above, are plotted, for the various categorizations in figure 1*A*–*D*. It can be seen that, for women with the same family history, the estimated probability in affected versus unaffected individuals was approximately four times greater for women in the youngest age category, but was only approximately two times greater for those in the oldest age category. That is, the effect of being affected depended on "age," a feature not discussed by Hartge et al. (1999).

From figure 1*A* and *B*, it can be seen that the probability approximately doubled for each affected first-degree relative. Figure 1*C* and *D* shows that, given a family history, the probability approximately doubled again if any of the affected relatives had an early diagnosis. These effects of family history were evident—and of the same magnitude on the log-odds scale—in both affected and unaffected individuals, another point not brought out by the CART analysis. We cannot, however, on the basis of the published data, test whether these effects were independent of one another.

In affected individuals, the probability approximately halved for each decade of age at *diagnosis*, whereas in unaffected individuals it approximately halved for every 2 decades of age at *testing*. The latter effect was not apparent from the CART analysis.

Thus, our analyses of the published Washington data suggest simple rules for how the probability that an Ashkenazi Jewish woman has inherited a founder mutation in BRCA1 or BRCA2 depends on her personal or family history of breast cancer. The effect of a having a single first-degree relative with breast cancer is generally a modest increase in prevalence equivalent to an odds ratio of either 2.7 from equation (1) or 2.2 from equation (2), both of which are similar to our estimate of 2.6 for a defined set of protein-truncating mutations in BRCA1 and BRCA2 that cause breast cancer in Australian women at age <40 years (Hopper et al. 1999). It would be interesting to do a similar analysis for a family history of ovarian cancer; table 1 of Hartge et al. (1999) suggests that having at least one first-degree relative with ovarian cancer is associated with an odds ratio of 4.1 (95% CI, 2.1–8.1) for having a founder mutation. Our analysis above has shown that this effect was no different from that of having a first-degree relative with breast cancer ($P = .3$).

Hartge et al. (1999) note that their observed probabilities for a given number of affected relatives were, in general, *less* than those derived from data collected in the setting of a cancer family clinic (Couch et al. 1997; Shattuck-Eidens et al. 1997). The discrepancy is likely to be due to the fact that the typical multiple-case families that come to such clinics also have cases with early-onset disease and other features—such as multiple primary cancers or breast and ovarian cancer in the same

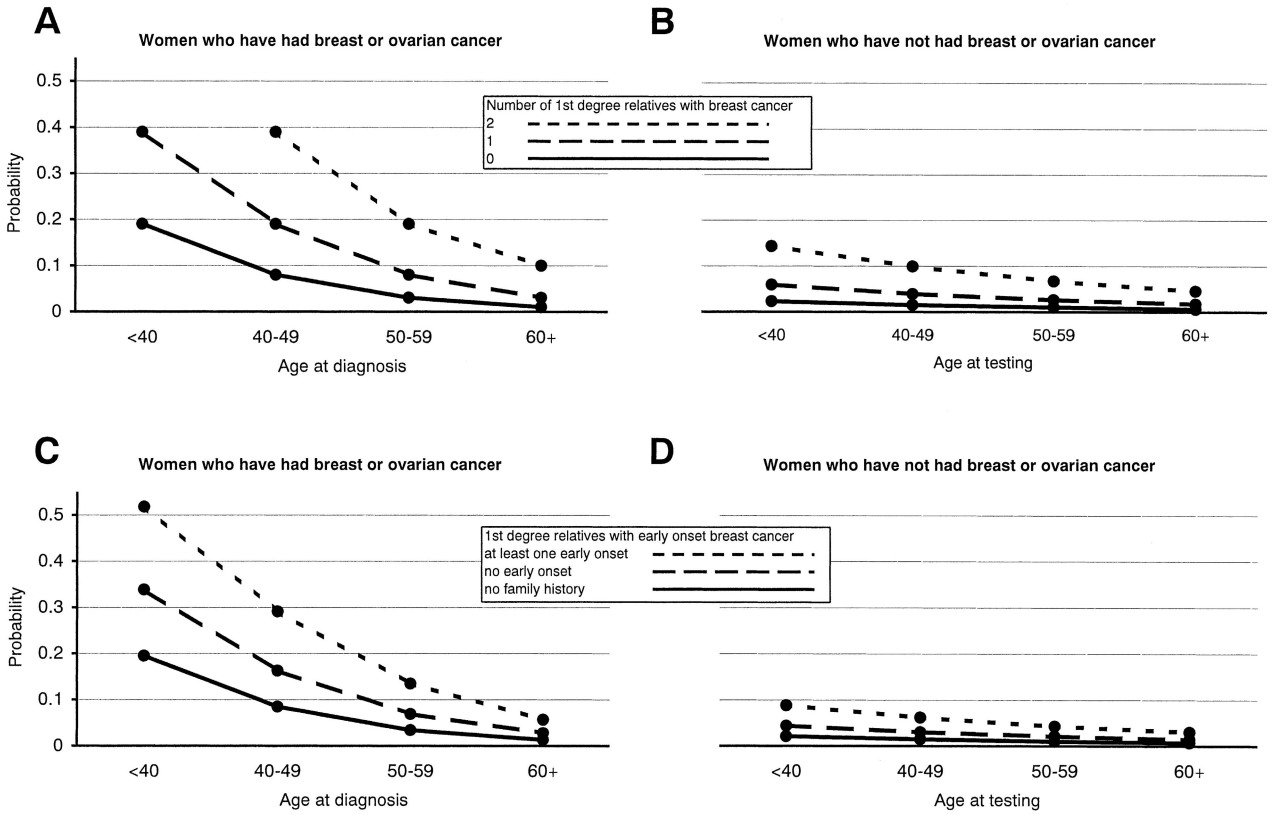

Figure 1 Estimated probability of carrying a founder mutation in BRCA1 or BRCA2 for Ashkenazi women. *A* and *C,* women who have had breast or ovarian cancer, as a function of age at diagnosis. *B* and *D,* women who have not had breast or ovarian cancer, as a function of age at testing. In *A* and *B,* family history is defined in terms of number of first-degree relatives with breast cancer; in *C* and *D,* family history is defined in terms of whether any first-degree relative had breast cancer at age <50 years (early diagnosis). (Note that because, in the data set, there were no women age <40 years who were both affected and who had more than one affected first-degree relative, a probability has not been assigned to that category.)

individual—that are also likely to increase their probability of being a carrier, as suggested by table 2 of Hartge et al. (1999). Without knowledge of *all* these predictors of mutation status it is not possible to make a valid inference from the clinic setting to the population.

The majority of "hereditary" cases of breast and ovarian cancer were "sporadic"; that is, 56% (15/27) of affected women who carried a founder mutation had "no breast cancer in [their] family" (see Hartge et al. 1999, table 4). The same can be considered to have been observed in Israel by Abeliovich et al. (1997), once attention is properly focused on their "unselected" sample of just 24 carriers, of whom 58% (14) had no affected first-degree relatives. This general observation also applies to population-based samples of non-Ashkenazi populations. For example, in the United Kingdom, Peto et al. (1999) found that of the 30 women who were diagnosed with breast cancer at age ≤45 years in whom a mutation in BRCA1 or BRCA2 was detected, 57% (17) had no family history of breast or ovarian cancer within three generations. In Australia, we have found that 72% (13/18) of BRCA1 or BRCA2 mutation-carrying cases diagnosed before age 40 years had no family history in the preceding two generations (Hopper et al. 1999).

Furthermore, only a small proportion of the typical women with "familial" breast cancer, defined as being affected *and* as having a family history of breast cancer, appear to have "hereditary" breast cancer, even in Ashkenazi Jews, in whom >2% of the population carries a mutation associated with a not-inconsiderable lifetime risk of ~50% (Struewing et al. 1997). The Washington study, in which there was a small bias toward living subjects having an excess of family history of breast cancer (Struewing et al. 1997), found that <10% of familial cases carried a founder mutation (Hartge et al. 1999), although it is conceivable that a smaller proportion may carry another "high-risk" mutation, either in BRCA1 or BRCA2 or, perhaps, another such gene whose normal function is to protect women from breast cancer. In our Australian population-based study of women with breast cancer at age <40 years, we detected a protein-truncating mutation in only 5% (6/120) of cases

with at least one affected first- or second-degree relative and in 9% (5/53) of cases with an affected first-degree relative, in a mutation screen covering between two-thirds and three-quarters of the coding regions of these two genes (Hopper et al. 1999, table 3).

Finally, the strong and highly significant effect that having had breast cancer has on the probability of being a mutation carrier could be used to derive an estimate of penetrance (i.e., age-specific cumulative risk of breast cancer), by using a case-control argument and appropriate population incidence rates and by taking into account the strong dependence of this effect on age, in which the odds ratio decreases from 10- to 2-fold across the four categories. In this regard, it is of interest that we found an average odds ratio of 9-fold for a set of protein-truncating mutations in BRCA1 and BRCA2 that cause early-onset breast cancer—and that this translates into a penetrance, until age 70 years, of just 40% when applied to Australian population rates (Hopper et al. 1999). Therefore, it is likely that a similar lifetime-penetrance estimate would apply to the founder mutations among U.S. Ashkenazi women, once the diminishing effect with age observed here has been counterbalanced by the ~30% higher underlying rates in the United States compared with Australia. Thus, population-based data on mutation carriers, such as those provided in some detail by Hartge and colleagues, are providing a new perspective on how genetic factors are evident in common diseases, challenging previous beliefs and language based on "monogenic" diseases (see Hopper et al. 1999).

JOHN L. HOPPER AND MARK A. JENKINS
Centre for Genetic Epidemiology, The University of Melbourne, Carlton, Victoria, Australia

Electronic-Database Information

Accession numbers and URL for data in this article are as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for BRCA1 [MIM 113705] and BRCA2 [MIM 600185])

References

Abeliovich D, Kaduri L, Lerer I, Weinberg N, Amir G, Sagi M, Zlotogora J, et al (1997) The founder mutations 185delAG and 5382insC in BRCA1 and 6174delT in BRCA2 appear in 60% of ovarian cancer and 30% of early-onset breast cancer patients among Ashkenazi women. Am J Hum Genet 60:505–514

Baker RJ, Nelder JA (1978) The GLIM system, release 3: generalized linear interactive modelling. Numerical Algorithms Group, Oxford

Couch FJ, DeShano ML, Blackwood MA, Calzone K, Stopfer J, Campeau L, Ganguly A, et al (1997) BRCA1 mutation in women attending clinics that evaluate the risk of breast cancer. N Engl J Med 336:1409–1415

Hartge P, Struewing JP, Wacholder S, Brody LC, Tucker, MA (1999) The prevalence of common BRCA1 and BRCA2 mutations among Ashkenazi Jews. Am J Hum Genet 64: 963–970

Hopper JL, Southey MC, Dite GS, Jolley DJ, Giles GG, McCredie MRE, Easton DF,and the Australian Breast Cancer Family Study (1999) 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 Biomarker Prevent 8:741–747

McCullagh P, Nelder JA (1983) Generalized linear models. Chapman & Hall, London

Peto J, Collins N, Barfoot R, Seal S, Warren W, Rahman N, Easton DF, et al (1999) Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer. J Natl Cancer Inst 91:943–949

Shattuck-Eidens D, Oliphant A, McClure M, McBride C, Gupte J, Rubano T, Pruss D, et al (1997) BRCA1 sequence analysis in women at high risk for susceptibility mutations: risk factor analysis and implications for genetic testing. JAMA 278:1242–1250

Struewing JP, Hartge P, Wacholder S, Baker SM, Berlin M, McAdams M, Timmerman MM, et al (1997) The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. N Engl J Med 336: 1401–1408

Address for correspondence and reprints: Dr. J. L. Hopper, The University of Melbourne, Centre for Genetic Epidemiology, 200 Berkeley Street, Carlton, Victoria 3053, Australia. E-mail: j.hopper@gpph.unimelb.edu.au

0002-9297/1999/6506-0033$02.00