ORIGINAL ARTICLE

Pre-test prediction models of *BRCA1* or *BRCA2* mutation in breast/ovarian families attending familial cancer clinics

M de la Hoya, O Díez, P Pérez-Segura, J Godino, J M Fernández, J Sanz, C Alonso, M Baiget, E Díaz-Rubio, T Caldés

J Med Genet 2003;**40**:503–510

See end of article for authors' affiliations

Correspondence to: Dr T Caldés, Laboratory of Molecular Oncology, Hospital Universitario San Carlos, c/Martín Lagos s/n, 28040 Madrid, Spain; tcaldes@hcsc.es

Revised version received 9 April 2003
Accepted for publication 10 April 2003

Objective: To test whether statistical models developed to calculate pre-test probability of being a *BRCA1/2* carrier can differentiate better between the breast/ovarian families to be referred to the DNA test laboratory.

Study design: A retrospective analysis was performed in 109 Spanish breast/ovarian families previously screened for germline mutations in both the *BRCA1* and *BRCA2* genes. Four easy to use logistic regression models originally developed in Spanish (HCSC model), Dutch (LUMC model), Finnish (HUCH model), and North American (U Penn model) families and one model based on empirical data of Frank 2002 were tested. A risk counsellor was asked to assign a subjective pre-test probability for each family. Sensitivity, specificity, negative and positive predictive values, and areas under receiver operator characteristics (ROC) curves were calculated in each case. Correlation between predicted probability and mutation prevalence was tested. All statistical tests were two sided.

Results: Overall, the models performed well, improving the performances of a genetic counsellor. The median ROC curve area was 0.80 (range 0.77-0.82). At 100% sensitivity, the median specificity was 30% (range 25-33%). At 92% sensitivity, the median specificity was 42% (range 33.3-54.2%) and the median negative predictive value was 93% (range 89.7-98%). *BRCA1* families tended to score higher risk than *BRCA2* families in all models tested.

Conclusions: All models increased the discrimination power of an experienced risk counsellor, suggesting that their use is valuable in the context of clinical counselling and genetic testing to optimise selection of patients for screening and allowing for more focused management. Models developed in different ethnic populations performed similarly well in a Spanish series of families, suggesting that models targeted to specific populations may not be necessary in all cases. Carrier probability as predicted by the models is consistent with actual prevalence, although in general models tend to underestimate it. Our study suggests that these models may perform differently in populations with a high prevalence of *BRCA2* mutations.

The identification of the breast cancer susceptibility genes *BRCA1* and *BRCA2* in the past decade[1 2] has permitted identification of presymptomatic subjects at risk of developing breast/ovarian cancer by means of a genetic test. Nowadays, many families with a moderate history of breast and/or ovarian cancer are self or physician referred to familial cancer clinics where genetic testing of these susceptibility genes is available. Unfortunately, the analysis is costly and time consuming and can cause considerable stress to many families. Moreover, a negative result does not imply a clear benefit, either psychological or clinical, given that genetic susceptibility cannot be ruled out in these families and other breast cancer genes unidentified to date may be involved.[3] Accordingly, it would be advantageous to target the available resources to test families with the highest probability of being mutation carriers. Thus, the development of an accurate pre-test determination of carrier probability has become in recent years a major topic in familial cancer clinics throughout the world.

In a 1996 policy statement, the American Society of Clinical Oncology (ASCO) suggested that gene mutation testing should be limited to subjects whose probability of carrying a mutation exceeds 10%.[4] There are a number of statistical approaches to calculating the pre-test probability of carrying a mutation.[5–11] However, subjective assessment by professional risk counsellors remains essential. Indeed, many familial cancer clinics do not refer families to the DNA laboratory in accordance with a calculated pre-test probability but establish "minimal entry criteria" (in terms of cancer phenotype) which all families selected for genetic testing must meet. Although no consensus exists, most familial cancer clinics will agree to select families with at least three cases of breast/ovarian cancer for genetic testing.[5–7]

The present study is focused on this type of family, which should be considered as high risk. However, only 30% of these families harbour a pathogenic mutation.[5–7] Therefore, the majority of the families currently referred to the DNA laboratory in cancer clinics throughout the world do not obtain any benefit from genetic testing. To reduce this proportion, a better understanding of the cancer phenotype associated with germline mutations in these genes is necessary.

Recently, some easy to use logistic regression models to calculate the pre-test probability that a family with a given cancer phenotype carries a *BRCA* mutation have been developed.[9 12–14] In most cases, these models have been devised with high risk families commonly attending familial cancer clinics. They take into account both *BRCA1* and *BRCA2* mutation status (with the exception of the model of Couch *et al*,[9] which is restricted to *BRCA1*), and make no assumption regarding prevalence or penetrance of these alleles in the target population. The performance of these models in an independent cohort of high risk families and their use in familial cancer clinics to reduce the number of *BRCA* negative families referred to the DNA laboratory have not been properly

Table 1 Characteristics of the sample

Characteristic	Mutation negative	Mutation positive	OR	p
No of pedigrees	72	37		
Proportion of pedigree with (95% CI)				
At least 1 ovarian cancer	0.21 (0.13 to 0.32)	0.57 (0.41 to 0.72)	4.99	0.0003
At least 1 breast and ovarian cancer in the same person	0.05 (0.02 to 0.13)	0.22 (0.12 to 0.38)	4.69	0.02
At least 1 bilateral breast cancer	0.18 (0.11 to 0.28)	0.32 (0.19 to 0.48)	2.18	0.09
At least 1 male breast cancer	0.04 (0.01 to 0.11)	0.13 (0.06 to 0.27)	3.59	0.1
Only unilateral breast cancer*	0.63 (0.51 to 0.73)	0.24 (0.13 to 0.40)	0.16	0.0001
Mean age at breast cancer diagnosis (95% CI)	49.6 (47.7 to 51.5)	43.6 (41.73 to 45.47)		0.0005
Mean age at diagnosis of the youngest breast cancer case (95% CI)	39.6 (37.7 to 41.47)	37.08 (34.30 to 39.84)		0.13

*Families with neither bilateral breast cancer nor male breast cancer or ovarian cancer.

evaluated. Moreover, these models have been developed in very specific populations and the predictive variables they use are similar but not identical, so it is not clear whether they can be implemented in populations other than the one for which they were devised.

The aim of our study was, therefore, to test the performance of easy to use prior probability models to decrease the number of true negative families that are currently referred to the DNA laboratory.

FAMILIES AND METHODS

We conducted this study in a clinic based cohort of 109 families. The Oncology and Genetics Departments of the Hospital de la Santa Creu i Sant Pau (Barcelona, Spain) and the Laboratory of Molecular Oncology, Department of Clinical Oncology, Hospital Clínico San Carlos (Madrid, Spain) submitted, respectively, 80 and 29 pedigrees and corresponding *BRCA1* and *BRCA2* results. These pedigrees had already been selected for complete *BRCA* gene sequencing on the basis of cancer family history information suggestive of an inherited breast and ovarian cancer predisposition (all pedigrees included at least three or more first or second degree relatives affected with breast or ovarian cancer in the same lineage). Pedigrees were constructed on the basis of an index case considered to have the highest probability of being a deleterious mutation carrier (generally the youngest affected subject available in each family). To construct pedigrees, patients were interviewed about their family history of cancer for information on cancer profiles and dates of diagnoses of all subjects, including first and second degree relatives of the index case. Characteristics of the study sample are summarised in table 1. Mutation analysis was performed in all index cases by either a combination of SSCP and PTT (Hospital de la Santa Creu i Sant Pau) or DGGE (Hospital Clínico San Carlos). In both cases, mutation screening protocols included all coding sequences and intron/exon boundaries.[12 15 16] The probability of carrying a *BRCA* mutation was calculated in each pedigree according to the four logistic regression models tested in this study. The model developed at the Hospital Clínico San Carlos (HCSC model)[12] considers as predictor variables the number of ovarian cancer cases in the family, mean age at diagnosis of breast cancer, and the presence/absence of concomitant breast and ovarian cancer in a single woman, bilateral breast cancer, and/or male breast cancer. The model of Peelen *et al*[14] was developed at the Leiden University Medical Centre (LUMC model). In this case, the predictor variables are the number of ovarian cancer cases in the family, the number of breast cancer cases in the family, mean age at diagnosis of breast cancer, and the presence/absence of bilateral breast cancer. The model of Vahteristo *et al*[13] was developed at the Helsinki University Central Hospital (HUCH model). The number of ovarian cancer cases in the family and the age of the youngest breast cancer patient in each family are the only predictor variables considered in this case. The model of Couch *et al*[9] was developed at the University of Pennsylvania (U Penn model). Predictor factors included average age at breast cancer diagnosis in the family under than 55 years, ovarian cancer in the family (particularly in a subject with breast cancer), and Ashkenazi Jewish ancestry. Data concerning the predictor variables of each model were available in all 109 pedigrees included in this analysis. We also calculated the probability of founding mutations according to the model of Frank 2002.[10] This is an empirical model which correlates prevalence of mutations in *BRCA1* and *BRCA2* with personal and family history of cancer. It is based on data from 10 000 subjects tested through Myriad Genetics. Probabilities according to the HCSC, LUMC, and HUCH models were calculated in a convenient Microsoft Excel format. In the case of the U Penn model, the probability for each possible permutation of the predictor variables has been previously calculated and tabulated[9] and we used these tabulated data (non-Ashkenazi heritage subset) to assign a probability to every family in our study sample. In the case of the model of Frank 2002,[10] we assigned a probability to each family according to the correlation found in 4716 non-Ashkenazi subjects (table 2).

An experienced risk counsellor from a familial cancer clinic was asked to evaluate each pedigree and assign a subjective pre-test probability for each family. Fifty percent of the risk counsellor practice was devoted specifically to breast-ovarian cancer susceptibility counselling. For the last five years, he has been counselling 8-10 Spanish breast-ovarian families each month. The risk counsellor was provided with data corresponding to the predictor variables used in the models. His assessment was based solely on his previous experience with Spanish breast/ovarian families and was not assisted by any pre-test probability statistical model.

Table 2 Relevant data concerning pre-test probability models

Entry criteria	No of pedigrees	*BRCA1* mutation prevalence	*BRCA2* mutation prevalence	Population	Designation	Reference
At least 3 breast/ovarian cancer	102	18%	12%	Spanish	HCSC	(12)
At least 3 breast/ovarian cancer	148	10.8%	8.8%	Finnish	HUCH	(13)
At least 3 breast/ovarian cancer	164	13.4%	7.3%	Dutch	LUMC	(14)
Women attending clinics	169	16%	Not analysed	White North-American	U Penn	(9)
Not applicable	4716	9.2%	5.9%	White North-American	Frank 2002	(10)

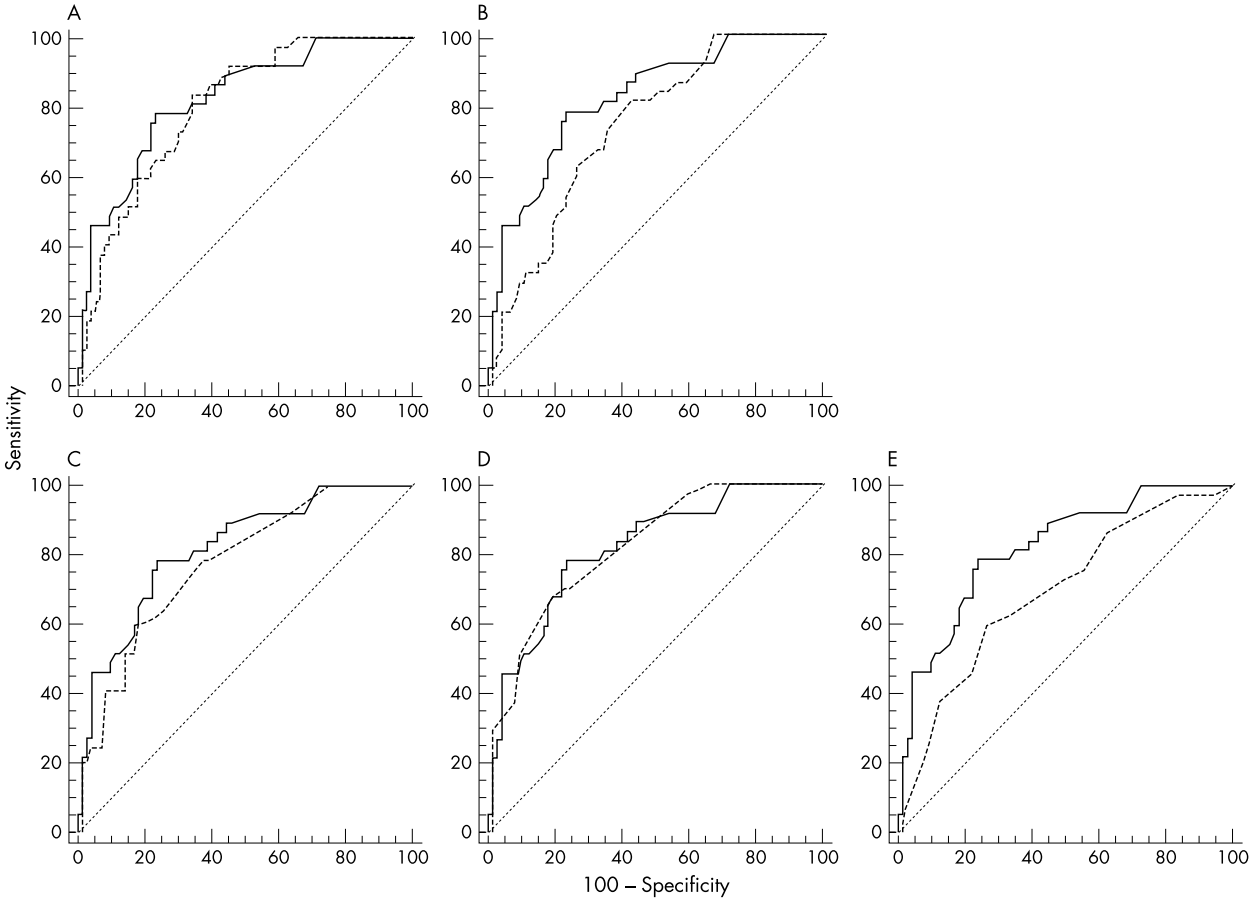

Figure 1 Receiver Operator Characteristic (ROC) curves. Comparison of HCSH (solid line with (A) LUMC, (B) HUCH, (C) U Penn, (D) Frank 2002, and (E) risk counsellor assessment.

Statistical analyses were two sided. Categorical variables were compared by the chi-square test and numerical variables by the *t* test. Receiver operator characteristics (ROC) curve areas, sensitivity, specificity, negative predictive values (PV−), positive predictive values (PV+), and correlation coefficients were calculated with the MedCalc software package.

RESULTS

The prevalence of *BRCA* mutations in our study sample was 33.9% (95% CI 26 to 43). This prevalence is consistent with that previously reported in Spanish and other populations.[5–6 9 12 15–17] Nineteen families carried a *BRCA1* mutation and 18 families had a *BRCA2* mutation. All mutations are predicted to produce a truncated protein and are considered pathogenic in the BIC database.[18] The spectrum of mutations is available on request from the authors. Other characteristics of the study sample are listed in table 1.

Families with ovarian cancer (57% *v* 21%, p=0.0003), concomitant breast and ovarian cancer in a single woman (22% *v* 5%, p=0.02), bilateral breast cancer (32% *v* 18%, p=0.09), and male breast cancer (13% *v* 4%, p=0.1) were more frequent in the *BRCA* positive group. However, only in the case of ovarian cancer and concomitant breast and ovarian cancer in a single woman did these differences reach statistical significance. Male breast cancer was the only phenotype clearly associated with *BRCA2* but not with *BRCA1* families (27.7% *v* 0%, p=0.02). The mean age at breast cancer diagnosis among women from mutation carrier families was lower than that for women from non-carrier families (43.6 years *v* 49.6 years, p=0.005). The mean age at diagnosis of the youngest breast cancer patient (relevant for the HUCH model) was also lower in *BRCA* positive families (37.8 years *v* 39.6 years) although the difference was not statistically significant. Overall, our study sample appears to be representative of breast/ovarian families commonly seen in familial cancer clinics and, therefore, relevant to our analysis.

As might be expected, the average pre-test probability of carrying a mutation was higher in positive than in negative families. Differences were as follows: 0.413 *v* 0.174 in the HCSC model, 0.356 *v* 0.133 in the LUMC model, 0.272 *v* 0.102 in the U Penn model, 0.381 *v* 0.174 in the HUCH model, 0.42 *v* 0.21 according to Frank 2002, and 0.58 *v* 0.46 according to the risk counsellor (all differences statistically significant at the p<0.001 level).

To compare the performance of the four logistic regression models with that of the Frank 2002 empirical data and the risk counsellor assessment, we calculated the Receiver Operator Characteristic (ROC) curve area, sensitivity, specificity, PV−, PV+, and the best discriminating probability threshold in each case (fig 1, table 3). The area under the ROC curve (a measure of the overall discrimination between BRCA positive and negative families) was 0.82 in the HCSC model, 0.80 in the LUMC model, 0.77 in the U Penn model, 0.77 in the HUCH model, 0.82 with the Frank 2002 prevalence, and 0.69 for the risk counsellor. Among statistical models, the maximum difference between the ROC areas (HCSC *v* HUCH, fig 1B) did not attain statistical significance (0.049 95% CI 0.05 to 0.148), indicating that these models have a similar power of discrimination. However, the ROC area calculated with data from the risk counsellor assessment was clearly lower when compared with any model. The maximum difference was 0.127 (95% CI 0.023 to 0.230) (HCSC *v* risk counsellor, fig 1D). This difference was significant (p=0.016) indicating that discrimination can be improved by using statistical models.

Table 3 Performance measures for five probabilistic models and one genetic counsellor in breast/ovarian Spanish pedigrees

Performances	HCSC	LUMC	U Penn	HUCH	Frank 2002	Risk counsellor
All pedigrees (n=109)						
ROC curve area (95% CI)	0.82 (0.73 to 0.88)	0.80 (0.72 to 0.88)	0.77 (0.68 to 0.85)	0.77 (0.69 to 0.84)	0.82 (0.73 to 0.89)	0.69 (0.60 to 0.78)
100% Sensitivity:						
Probability threshold	>0.089	>0.035	>0.021	>0.047	>0.11	>0.15
Specificity (95% CI)	27.8% (17.9 to 39.6)	32% (21.4 to 44.0)	25% (15.5 to 36.6)	31.9% (21.4 to 44.0)	33.3% (22.7 to 45.4)	0%
PV^-	100%	100%	100%	100%	100%	–
PV^+	41.6%	43%	40.7%	43%	43.5%	33.9%
92% sensitivity						
Probability threshold	>0.114	>0.075	>0.032	>0.054	>0.175	>0.30
Specificity (95% CI)	45.8% (34.0 to 58.0)	54.2% (42.0 to 66.0)	36.1% (25.1 to 48.3)	33.3% (22.7 to 45.4)	40.3% (28.9 to 52.5)	26.4% (16.7 to 38.1)
PV^-	91.7%	92.9%	89.7%	92.3%	96.7%	86.4%
PV^+	46.6%	50.7%	42.5%	42.2%	45.6%	39.1%
Optimal from a statistical point of view						
Probability threshold	>0.232	>0.094	>0.117	>0.148	>0.341	>0.55
Sensitivity	78.4% (61.8 to 90.1)	86.5% (71.2 to 95.4)	59.5% (42.1 to 75.2)	73.0% (55.9 to 86.2)	67.6% (50.2 to 82.0)	59.5% (42.1 to 75.2)
Specificity	76.4% (64.9 to 85.6)	63.9% (51.7 to 74.9)	81.9% (71.1 to 90.0)	70.8% (58.9 to 81.0)	80.6% (69.5 to 88.9)	73.6% (61.9 to 83.3)
PV^-	87.3%	90.2%	79.7%	83.6%	82.9%	77.9
PV^+	63%	55.2%	62.9%	56.2%	64.1%	53.7
Breast only pedigrees (n=72)						
ROC curve area (95% CI)	0.78 (0.66 to 0.87)	0.77 (0.66 to 0.86)	0.72 (0.60 to 0.82)	0.73 (0.61 to 0.82)	0.74 (0.62 to 0.84)	0.67 (0.55 to 0.78)
100% Sensitivity						
Probability threshold	>0.089	>0.035	>0.021	>0.047	>0.11	>0.25
Specificity (95% CI)	33.9% (21.8 to 47.8)	37.5% (24.9 to 51.5)	32.1% (20.3 to 46.0)	39.3% (26.5 to 53.2)	41.1% (28.1 to 55)	17.9% (8.9 to 30.4)
PV^-	100%	100%	100%	100%	100%	100%
PV^+	30.2%	31.4%	29.6%	32%	32.7%	25.8%

Three relevant probability thresholds were selected for our analysis: the 100% sensitivity threshold, the 92% sensitivity threshold (which we consider acceptable in clinical practice), and the best discriminating probability threshold (best performance from a statistical point of view). It is interesting to note that the best discriminating probability threshold was not clinically relevant in any statistical model. This was because the sensitivities were well below 90% in all cases, ranging from 86.5% (LUMC model) to 59.5% (U Penn model). Similarly, the sensitivity reached by the risk counsellor (59.5%) did not have any clinical relevance. However, the specificity ranged from 25% (U Penn model) to 33% (Frank 2002) if families with pre-test probabilities above the 100% sensitivity threshold were selected for testing. To compare the performances of these models further, we chose an arbitrary but clinically acceptable 92% sensitivity threshold (table 3). By selecting families with probabilities above this threshold, specificity ranged from 33.3% (HUCH model) to 54.2% (LUMC model). The risk counsellor specificity was clearly lower (26.4%), although the difference only reached statistical significance when compared with the HCSC ($p<0.05$) and LUMC models ($p<0.05$). Overall, the data shown in table 3 indicate that the selection of a suitable pre-test probability threshold (which is different in each model) will better differentiate the families to be referred to the DNA laboratory. Ovarian cancer is an important variable in all predictor models. It might therefore be possible that the performance of these models varies with the proportion of breast/ovarian families present in the cohort. To test this hypothesis, we performed a subanalysis in the 72 breast only families (families with no single case of ovarian cancer reported) present in our cohort. This subset of families included four *BRCA1* families, 12 *BRCA2* families, and 56 negative families. As shown in table 3, with 100% sensitivity, neither ROC area nor specificity are severely impacted, suggesting that these models are not dependent on ovarian cancer to discriminate *BRCA* positive from *BRCA* negative families.

To analyse these models further, it is interesting to study the characteristics (if any) of true positive families which tend to be misclassified as negatives. In our study series (37 *BRCA*

Table 4 Performances in *BRCA1* families *v BRCA2* families

	Average probability			ROC area		
Model	*BRCA1* families (n=19)	*BRCA2* families (n=18)	p	*BRCA2* negative set (n=91)	*BRCA1* negative set (n=90)	p
HCSC	0.466	0.356	NS	0.84 (0.74 to 0.90)	0.80 (0.70 to 0.88)	NS
LUMC	0.47	0.236	0.006	0.86 (0.77 to 0.92)	0.75 (0.65 to 0.84)	<0.05
U Penn	0.36	0.177	<0.001	0.83 (0.74 to 0.90)	0.71 (0.60 to 0.80)	<0.05
HUCH	0.50	0.254	0.004	0.85 (0.76 to 0.92)	0.68 (0.58 to 0.78)	<0.05
Frank 2002	0.51	0.33	0.005	0.89 (0.81 to 0.95)	0.74 (0.61 to 0.80)	<0.05
Risk counsellor	0.56	0.59	NS	0.67 (0.56 to 0.76)	0.71 (0.61 to 0.80)	NS

NS, not significant.

Table 5 Correlation between pre-test probabilities and mutation prevalence

Model	Pre-test probability quartiles				Correlation coeficient
	1st (n=28)	2nd (n=28)	3rd (n=28)	4th (n=25)	
HCSC					
Average pre-test probability	0.07	0.08	0.27	0.60	0.994
Prevalence of mutations (95% CI)	0.10 (0.03 to 0.26)	0.17 (0.07 to 0.35)	0.39 (0.23 to 0.57)	0.72 (0.54 to 0.85)	p=0.006
LUMC					
Average pre-test probability	0.02	0.07	0.21	0.57	0.947
Prevalence of mutations (95% CI)	0.03 (0.0 to 0.17)	0.25 (0.13 to 0.43)	0.43 (0.27 to 0.61)	0.68 (0.49 to 0.82)	p=NS
U Penn					
Average pre-test probability	0.02	0.05	0.12	0.47	0.869
Prevalence of mutations (95% CI)	0.11 (0.04 to 0.28)	0.21 (0.10 to 0.39)	0.46 (0.29 to 0.64)	0.60 (0.42 to 0.76)	p=NS
HUCH					
Average pre-test probability	0.03	0.10	0.24	0.65	0.944
Prevalence of mutations (95% CI)	0.07 (0.02 to 0.22)	0.25 (0.13 to 0.43)	0.43 (0.27 to 0.61)	0.64 (0.46 to 0.79)	p=NS
Frank 2002					
Average pre-test probability	0.09	0.18	0.30	0.57	0.933
Prevalence of mutations (95% CI)	0 (0 to 0.12)	0.07 (0.02 to 0.22)	0.57 (0.39 to 0.73)	0.76 (0.57 to 89)	p=NS
Risk counsellor					
Average pre-test probability	0.27	0.44	0.58	0.74	0.992
Prevalence of mutations (95% CI)	0.14 (0.06 to 0.31)	0.25 (0.13 to 0.43)	0.43 (0.27 to 0.61)	0.56 (0.38 to 0.73)	p=0.007

positive families) a significant concordance among models was observed. The HCSC, LUMC, and U Penn models misclassified the SP18 (*BRCA2*), SP122 (*BRCA2*), and SP46 (*BRCA1*) families as negatives by using the 92% sensitivity threshold. The same families are misclassified if Frank 2002 tables are used. These families shared a common phenotype: three unilateral breast cancer cases with a median age of 50 years at diagnosis (borderline minimal entry criteria). The HUCH model, which considers the youngest age but not the median age at diagnosis, ruled out the SP46 family but correctly selected the SP18 and SP122 families. Both families include a breast cancer case diagnosed at the age of 33. By contrast, the SP33 (with three unilateral breast cancer cases, one bilateral breast cancer case, and a median age of 46 at diagnosis) and the SC182 (with five breast cancer cases with a median age of 46.8 at diagnosis) *BRCA2* families were selected for analysis by the HCSC, LUMC, and U Penn model and by Frank 2002, but ruled out by the HUCH model. The median age at diagnosis is low (46 years) in both families, whereas the youngest age at diagnosis is not especially low (44 years).

Interestingly, four out of five positive families misclassified as negative by at least one of these models are related to *BRCA2*, suggesting that the discrimination power of these models is lower in these families. Indeed, the average probabilities scored by the *BRCA1* families with the LUMC, HUCH, and U Penn models were twice as high as those scored by the *BRCA2* families. A similar difference was observed with Frank 2002. The differences were smaller and not statistically significant for the HCSC model. The probabilities were almost identical in the *BRCA1* and *BRCA2* families based on the risk counsellor assessment (table 4).

These data suggest that the presence of *BRCA2* families in the study sample impairs the discrimination power of the probabilistic models. To test this hypothesis, we calculated ROC curves in alternative study samples from which the *BRCA1* or *BRCA2* related families were selectively removed (*BRCA1* negative and *BRCA2* negative study sample, respectively). The ROC curve areas calculated with the *BRCA2* negative study sample were higher than those obtained with the *BRCA1* negative study sample and than those calculated with the original study sample in the five models (table 4). However, in the case of the HCSC, these differences were modest. Interestingly, of the 10 positive families with the lowest pre-test probabilities, five families were *BRCA2* related in

Table 6 Correlation between pre-test probabilities and mutation prevalence in breast only families

Model	Pre-test probability quartiles				Correlation coeficient
	1st (n=18)	2nd (n=18)	3rd (n=18)	4th (n=18)	
HCSC					
Average pre-test probability	0.05	0.10	0.15	0.40	0.977
Prevalence of mutations (95% CI)	0 (0.0 to 0.18)	0.16 (0.05 to 0.38)	0.22 (0.09 to 0.45)	0.50 (0.29 to 0.71)	p=0.023
LUMC					
Average pre-test probability	0.01	0.05	0.09	0.26	0.829
Prevalence of mutations (95% CI)	0 (0.0 to 0.18)	0.16 (0.05 to 0.38)	0.33 (0.16 to 0.56)	0.38 (0.20 to 0.61)	p=NS
U Penn					
Average pre-test probability	0.02	0.04	0.05	0.10	0.86
Prevalence of mutations (95% CI)	0 (0.0 to 0.18)	0.16 (0.05 to 0.38)	0.33 (0.16 to 0.56)	0.38 (0.20 to 0.61)	p=NS
HUCH					
Average pre-test probability	0.02	0.06	0.11	0.23	0.86
Prevalence of mutations (95% CI)	0 (0.0 to 0.18)	0.27 (0.12 to 0.50)	0.16 (0.05 to 0.38)	0.44 (0.24 to 0.66)	p=NS
Frank 2002					
Average pre-test probability	0.08	0.16	0.18	0.33	0.97
Prevalence of mutations (95% CI)	0 (0.0 to 0.18)	0.11 (0.03 to 0.33)	0.28 (0.13 to 0.51)	0.50 (0.29 to 0.71)	p=0.03
Risk counsellor					
Average pre-test probability	0.24	0.37	0.49	0.67	0.81
Prevalence of mutations (95% CI)	0.06 (0.01 to 0.26)	0.28 (0.13 to 0.51)	0.16 (0.05 to 0.38)	0.39 (0.20 to 0.61)	p=NS

Table 7 Correlation between pre-test probabilities and mutation prevalence in breast/ovarian families

Model	Pre-test probability quartiles				Correlation coeficient
	1st (n=10)	2nd (n=9)	3rd (n=9)	4th (n=9)	
HCSC					
Average pre-test probability	0.16	0.28	0.44	0.77	0.85
Prevalence of mutations (95% CI)	0.20 (0.06 to 0.51)	0.66 (0.35 to 0.88)	0.55 (0.28 to 0.81)	0.89 (0.57 to 0.98)	p=NS
LUMC					
Average pre-test probability	0.12	0.27	0.48	0.81	0.72
Prevalence of mutations (95% CI)	0.30 (0.11 to 0.60)	0.55 (0.26 to 0.81)	0.77 (0.45 to 0.93)	0.66 (0.35 to 0.88)	p=NS
U Penn					
Average pre-test probability	0.14	0.26	0.40	0.69	0.88
Prevalence of mutations (95% CI)	40 (0.17 to 0.69)	0.55 (0.26 to 0.81)	0.44 (0.19 to 0.73)	0.89 (0.57 to 0.98)	p=NS
HUCH					
Average pre-test probability	0.22	0.41	0.60	0.87	0.50
Prevalence of mutations (95% CI)	30 (0.11 to 0.60)	0.55 (0.26 to 0.81)	0.88 (0.57 to 0.98)	0.55 (0.26 to 0.81)	p=NS
Frank 2002					
Average pre-test probability	0.21	0.41	0.51	0.72	0.933
Prevalence of mutations (95% CI)	0.10 (0.02 to 0.40)	0.66 (0.35 to 0.98)	0.66 (0.35 to 0.98)	0.88 (0.57 to 0.98)	p=NS
Risk counsellor					
Average pre-test probability	0.39	0.59	0.69	0.76	0.847
Prevalence of mutations (95% CI)	0.50 (0.24 to 0.76)	0.55 (0.26 to 0.81)	0.55 (0.26 to 0.81)	0.66 (0.35 to 0.88)	p=NS

the HCSC model, six in the LUMC model, seven in the U Penn model, eight in Frank 2002, and nine in the HUCH model. Taken together these data indicate that all models, but especially U Penn, Frank 2002, and HUCH, discriminate *BRCA1* better than *BRCA2* families. Interestingly, there is no similar bias when risk counsellor assessment is considered. The average probability was almost identical in *BRCA1* and *BRCA2* families (0.56 *v* 0.59). Moreover, the three families ruled out at 92% sensitivity threshold were *BRCA1* related, only four out of the 10 families with the lowest pre-test probability were *BRCA2* related, and the calculated ROC area was higher in the *BRCA1* negative than in the *BRCA2* negative sample (table 4).

The logistic regression models analysed in this study take into consideration cancer phenotype but not pedigree structure. Therefore, it may well be that although the pre-test probabilities calculated by the models are useful to discriminate positive families, they do not reflect true probabilities, and are therefore not good estimators of prevalence. To investigate the relationship between pre-test probability and the prevalence of mutations, we partitioned our data set into quartiles by pre-test probabilities, and the prevalence of mutations was calculated in each quartile (table 5). The sample size was 28 in the first three quartiles and 25 in the last one. The correlation coefficient between pre-test probabilities and prevalence after genetic testing was 0.994 for the HCSC model (p=0.006), 0.947 for the LUMC model (p=0.053), 0.944 for the HUCH model (p=0.056), 0.869 for the U Penn model (p=0.131), and 0.933 for Frank 2002 (p=0.06). These data indicate that a reasonable correlation between pre-test probabilities and mutation prevalence exists, which is the highest in the HCSC model. However, it is also clear that models tend to underestimate prevalence (LUMC and HUCH predictions are below the 95% interval in two quartiles, U Penn in three quartiles, and Frank 2002 in one quartile) and some corrections should be done in the models to fit pre-test probabilities and prevalence. It should be pointed out that the correlation obtained by the risk counsellor (table 5) was among the best, although in this case probabilities were not underestimated but clearly overestimated (approximately two-fold). To test if mutation prevalence was equally underestimated in breast only and breast/ovarian families, we performed a subanalysis of predicted probability/prevalence correlation in these two groups separately (tables 6 and 7). Taken together, the data indicate that all models tend to underestimate mutation prevalence both in breast only families and breast/ovarian families, although this trend is more evident in breast only families.

DISCUSSION

Since the cloning of the breast cancer susceptibility genes *BRCA1* and *BRCA2*,[1 2] a number of statistical models have been developed to predict best the pre-test probability of carrying a germline mutation in one of these genes. Most of these models have not been properly evaluated to date.

Several pre-test probability models do exist.[19] Among them, the models by Couch, Shattuck-Eidens, Frank, and BRCAPRO are widely used, although recently other models focused on different ethnic populations have been developed. Each model has been developed with different methodology, sample size, and population characteristics, and consequently each model has unique attributes and limitations, making them difficult to compare directly in a given set of families.

We have performed a retrospective analysis of easy to use statistical models predicting pre-test probability of carrying a *BRCA* mutation in a series of Spanish breast/ovarian families attending familial cancer clinics (all of them with cancer family history information suggestive of an inherited breast or breast/ovarian cancer predisposition). We did not pretend to test the sensitivity of these models at the lower end of the scale (our cohort did not include low risk families) but to test the performances of these models in high risk families who had already been selected for genetic testing on the basis of cancer family history. There are many probability models which can be investigated.[19] We have decided to test four easy to use models which have been originally developed in different ethnic populations (HCSC in Spanish, LUMC in Dutch, HUCH in Finnish, and U Penn in white North American populations) but share a number of characteristics: a logistic regression approach, almost identical entry criteria, and prediction of familial not individual risk. To the best of our knowledge, this is the first time that these models have been tested in an independent set of families. On the other hand, we have tested the performance of Frank 2002, which represents empirical data obtained in white North Americans and it is therefore an empirical and not model approach. However, as our main objective was to test easy to use models in high risk families, we decided not to include in our analysis two other widely used models, Shattuck-Eidens and BRCAPRO. The Shattuck-Eidens model is not applicable to women diagnosed with breast cancer under 30 and therefore it is not applicable to 11 families in our cohort. Moreover, this model is not appropriate for high risk families. On the other hand, the BRCAPRO is not an easy to use model and it has some practical limitations, specific computer software is required, and it is limited to first and second degree relatives. More importantly, in this study

we address familial risk while BRCAPRO gives a personal rather than a familial a priori probability and sometimes it is not obvious which proband to select to capture familial risk best.

Overall, our study shows no major differences in discrimination power (as measured by ROC areas) among the models. It should be noted that all models increased the discrimination power of an experienced risk counsellor, suggesting that their use is valuable in the context of clinical counselling and genetic testing to optimise selection of patients for screening. However, given that the ROC areas compare the performances of the models over the complete range of sensitivity, they do not accurately reflect the true merits in familial cancer clinics (as only the upper limit range of sensitivity is relevant in this case). In all the models tested, the optimal probability threshold is not clinically relevant (sensitivities are well below 90%), although it may be useful in some applications, for instance, in identifying a certain number of positive families with minimal screening effort. Our study indicates that these models can improve mutation risk assessment in high risk families commonly seen in a familial cancer clinic. For example, by calculating pre-test probability with the LUMC model and selecting for genetic testing those families scoring a probability greater than 3.5%, all the positive families would be selected (100% sensitivity) and as many as 32% of the true negative families could be considered as low risk families. Using the same model, by selecting all families scoring a probability higher than 7.5%, the sensitivity remains higher than 90%, and 54% of the true negative families could be considered low risk. Therefore, with this model, 41 out of 100 breast/ovarian families currently considered as high risk (38 negative families and three positive families) could be reconsidered as low risk. The HCSC model achieved similar performances. Recently, the performance of BRCAPRO was validated in 148 high risk families.[20] By using a >10 *BRCA* mutation probability, sensitivity was 92%, specificity 32%, and PV− 84%. These data, taken together with ours, suggest that although with different intrinsic characteristic, easy to use logistic regression models, computer assisted BRCAPRO, and empirical data on mutation prevalence, as exemplified by Frank 2002, may have similar performances in high risk breast/ovarian families.

Overall, we consider that the models tested in this study perform well. However, some differences are also observed. For instance, our data indicate that within the range of sensitivity required in a genetic cancer clinic, the specificity of U Penn and HUCH models are well below the range observed in the LUMC and HCSC models. At the same time, the *BRCA2* families tend to score lower pre-test probabilities than *BRCA1* families in all models, but this trend was more evident in the U Penn and HUCH models (table 4). This can be attributed to a strong bias of these models towards *BRCA1* families and explains why the overall performance in our sample test (*BRCA2* accounting for 50% of the positive families) was much better with the LUMC and HCSC models.

A *BRCA1* bias was expected in the U Penn model (as this model is only strictly applicable to *BRCA1* mutations)[9] but not in the other models. This could reflect a true milder *BRCA2* phenotype. This is in agreement with ovarian and breast cancer penetrance estimates of *BRCA2*, which are lower than those of *BRCA1*.[21] Therefore, a milder phenotype might be expected in these families, which is reflected by scoring lower pre-test probabilities (for instance, ovarian cancer which is very important in all models is less frequent in *BRCA2* families). These considerations raise the question of whether these models may be useful in populations with a high prevalence of *BRCA2* mutations. However, we have performed a subanalysis in breast only families (12 out of 16 positive families being *BRCA2* related), which suggests that *BRCA2* families do not impair the performance of models.

Screening protocols like DGGE, SSCP+PTT, or others are widely used for *BRCA1* and *BRCA2* testing, although they are not 100% sensitive.[22] A more sensitive protocol analysis (full sequencing plus gene rearrangement analysis) can be expected to increase slightly the prevalence of mutations in our group of families, probably increasing the performances of both models and risk counsellor assessment.

In conclusion, the four logistic regression models tested may be of use to familial cancer clinics although a *BRCA1* related bias was observed in all of them. Models developed in specific populations (such as Dutch or Finnish) can be used in other populations (Spanish in this case). Our data suggest that there is no need for population specific logistic models but rather a need for models based on larger sets of families (this may be more easily accomplished by pooling pedigrees from different populations). At present, pre-test probability models are not good enough to rule out families from genetic analysis solely on the basis of a pre-test probability threshold. However, these models can help a risk counsellor to estimate gene mutation probability in a more consistent way. This estimation is an important initial task for risk counsellors, allows for more focused management, and permits reduction of the number of families considered as high risk.

ACKNOWLEDGEMENT

This work was supported by Fondo de Investigación Sanitaria (FIS) grant number 01/3040, 01/0024-02, 01/0024-03. Javier Godino is a fellow of FIS (99/1906).

Authors' affiliations
M de la Hoya, J Godino, J M Fernández, T Caldés, Laboratory of Molecular Oncology, Hospital Clínico San Carlos, Madrid, Spain
O Díez, M Baiget, Department of Genetics, Hospital de la Santa Creu I Sant Pau, Barcelona, Spain
P Pérez-Segura, E Díaz-Rubio, Department of Clinical Oncology, Hospital Clínico San Carlos, Madrid, Spain
J Sanz, C Alonso, Department of Clinical Oncology, Hospital de la Santa Creu I Sant Pau, Barcelona, Spain

REFERENCES

1 **Miki Y**, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W, *et al*. A strong candidate for the breast cancer and ovarian cancer susceptibility gene BRCA1. *Science* 1994;**266**:66-71.
2 **Wooster R**, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G. Identification of the breast cancer susceptibility gene BRCA2 to chromosome 13q12. *Science* 1995;**265**:2088-90.
3 **Eng C**, Hampel H, de la Chapelle A. Genetic testing for cancer predisposition. *Annu Rev Med* 2000;**52**:371-400.
4 **National Comprehensive Cancer Network**. CCN practice guidelines: genetics/familial high risk cancer. *Oncology* 1999;**13**:161-96.
5 **Peelen T**, van Vliet M, Petrij-Bosch A, Mieremet R, Szabo C, van den Ouweland AM, Hogervorst F, Brohet R, Ligtenberg MJ, Teugels E, van der Luijt R, van der Hout AH, Gille JJ, Pals G, Jedema I, Olmer R, van Leeuwen I, Newman B, Plandsoen M, van der Est M, Brink G, Hageman S, Arts PJ, Bakker MM, Devilee P, *et al*. A high proportion of novel mutations in BRCA1 with strong founder effects among Dutch and Belgian hereditary breast and ovarian cancer families. *Am J Hum Genet* 1997;**60**:1041-9.
6 **Wagner TM**, Moslinger RA, Muhr D, Langbauer G, Hirtenlehner K, Concin H, Doeller W, Haid A, Lang AH, Mayer P, Ropp E, Kubista E, Amirimani B, Helbich T, Becherer A, Scheiner O, Breiteneder H, Borg A, Devilee P, Oefner P, Zielinski C. BRCA1-related breast cancer in Austrian breast and ovarian cancer families: specific BRCA1 mutation and pathological characteristics. *Int J Cancer* 1998;**77**:354-60.
7 **Osorio A**, Barroso A, Martinez B, Cebrian A, San Roman JM, Lobo F, Robledo M, Benitez J. Molecular analysis of the BRCA1 and BRCA2 genes in 32 breast and/or ovarian cancer Spanish families. *Br J Cancer* 2000;**82**:1266-70.
8 **Shattuck-Eidens D**, Oliphant A, McClure M, McBride C, Gupte J, Rubano T, Pruss D, Tavtigian SV, Teng DH, Adey N, Staebell M, Gumpper K, Lundstrom R, Hulick M, Kelly M, Holmen J, Lingenfelter B, Manley S, Fujimura F, Luce M, Ward B, Cannon-Albright L, Steele L, Offit K, Thomas A, *et al*. BRCA1 sequence analysis in woman at high risk for susceptibility mutations: risk factor analysis and implications for genetic testing. *JAMA* 1997;**278**:1242-50.
9 **Couch FJ**, DeShano ML, Blackwood MA, Calzone K, Stopfer J, Campeau L, Ganguly A, Rebbeck T, Weber BL. BRCA1 mutations in woman attending clinics that evaluated the risk of breast cancer. *N Engl J Med* 1997;**336**:1409-15.
10 **Frank TS**, Deffenbaugh AM, Reid JE, Hulick M, Ward BE, Lingenfelter B, Gumpper KL, Scholl T, Tavtigian SV, Pruss DR, Critchfield GC. Clinical

characteristics of individuals with germline mutations in BRCA1 and BRCA2: analysis of 10000 Individuals. *J Clin Oncol* 2002;**20**:1480-90.
11 **Parmigiani G**, Berry DA, Aguilar O. Determining carrier probabilities for breast cancer-susceptibility genes BRCA1 and BRCA2. *Am J Hum Genet* 1998;**62**:145-58.
12 **de la Hoya M**, Osorio A, Godino J, Sulleiro S, Tosar A, Perez-Segura P, Fernandez C, Rodriguez R, Diaz-Rubio E, Benitez J, Devilee P, Caldes T. Association between BRCA1 and BRCA2 mutations and cancer phenotype in Spanish breast/ovarian cancer families: implications for genetic testing. *Int J Cancer* 2002;**97**:466-71.
13 **Vahteristo P**, Eerola H, Tamminen A, Blomqvist C, Nevanlinna H. A probability model for predicting BRCA1 and BRCA2 mutations in breast and breast-ovarian cancer families. *Br J Cancer* 2001;**84**:704-8.
14 **Peelen T**. *Genetics of hereditary breast and ovarian cancer in the Netherlands*. PhD thesis, Leiden University, Leiden, The Netherlands, 2001:115-28.
15 **Campos B**, Diez O, Domenech M, Baena M, Pericay C, Balmana J, del Rio E, Sanz J, Alonso C, Baiget M. BRCA2 mutation analysis of 87 Spanish breast/ovarian cancer families. *Ann Oncol* 2001;**12**:1699-703.
16 **Diez O**, Cortes J, Domenech M, Brunet J, Del Rio E, Pericay C, Sanz J, Alonso C, Baiget M. BRCA1 mutation analysis in 83 Spanish breast and breast/ovarian cancer families. *Int J Cancer* 1999;**83**:465-9.
17 **German Consortium for Hereditary Breast cancer**. Comprehensive analysis of 989 patients with breast or ovarian cancer provides BRCA1 and BRCA2 mutation profiles and frequencies for the German population. *Int J Cancer 2002;***97**:472-80.
18 **Breast Cancer Informative Core Database at www**.nchgr.nih.gov/Intramural_Research/Lab_transfer/Bic/guidelines.html.
19 **Domchek S**, Eisen A, Calzone K, Stopfer J, Blackwood A, Weber BL. Application of breast cancer risk prediction models in clinical practice. *J Clin Oncol* 2003;**21**:593-601.
20 **Euhus DM**, Smith KC, Robinson L, Stucky A, Olopade OI, Cummings S, Garber JE, Chittenden A, Mills GB, Rieger P, Esserman L, Crawford B, Hughes KS, Roche CA, Ganz PA, Seldon J, Fabian CJ, Klemp J, Tomlinson G. Pretest prediction of BRCA1 or BRCA2 mutation by risk counselors and the computer model BRCAPRO. *J Natl Cancer Inst* 2002;**94**:844-51.
21 **Ford D**, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, Bishop DT, Weber B, Lenoir G, Chang-Claude J, Sobol H, Teare MD, Struewing J, Arason A, Scherneck S, Peto J, Rebbeck TR, Tonin P, Neuhausen S, Barkardottir R, Eyfjord J, Lynch H, Ponder BA, Gayther SA, Zelada-Hedman M, 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.
22 **Eng C**, Brody LC, Wagner T, Devilee P, Vijg J, Szabo C, Tavtigian SV, Nathanson KL, Ostrander E, Frank TS. Interpreting epidemiological research: blinded comparison of methods used to estimate the prevalence of inherited mutations in *BRCA1*. *J Med Genet* 2001;**38**:824-33.