

A METHOD FOR DETERMINING FAMILIAL CANCER RISKS IN CLINICAL PRACTICE

R.S. HOULSTON and D. FORD

Section of Epidemiology, Institute of Cancer Research, Sutton, Surrey, UK

SUMMARY

There is interest in estimating familial cancer risks in clinical practice for counselling and determining patients' screening requirements. Empiric methods can be used to estimate an individual's risk, however, every family history is unique making such methods relatively non-specific. In contrast if an underlying genetic model can be assumed the risk of disease can be calculated for any individual using his or her family history. A method of estimating familial cancer risks based on segregation models and linkage data is presented and its implementation discussed.

KEY WORDS Survival methods Segregation analysis Linkage analysis
Familial cancer risk

INTRODUCTION

There is considerable interest in estimating consultands' familial cancer risks for genetic counselling, and as a guide to determining entry into screening programmes and trials of chemoprevention. Empiric risks can be used in clinical practice for counselling, however, every family history is unique and in many circumstances an empiric risk may be non-specific or not available. In contrast if an underlying genetic model can reasonably be assumed (specifying mode of inheritance, gene frequency and age-specific risks) the risk of disease can be calculated for any individual using his or her family history.

Complex segregation analysis aims to define the most probable (if any) genetic mechanism involved in the aetiology of a disease. In essence this involves comparing the observed pattern of disease incidence in systematically ascertained pedigrees, given that certain individuals are known a-priori to be affected, with that predicted by a number of different models. The best fitting model is generally determined by maximum likelihood (Weiss, 1993). Once a causative gene has been identified specific estimates of the gene frequency and risks conferred by a mutation may become available through linkage studies and/or follow-up of gene carriers.

This article provides a method for estimating familial cancer risks based on the results of segregation and linkage data for use in clinical practice.

Correspondence to: R.S. Houlston, Section of Epidemiology, Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK. Tel 0181-643-8901. Fax 0181-770-7876. e-mail R.Houlston@ICR.AC.UK.

*Received 8 March, 1996
Revised 28 March, 1996*

METHODOLOGY

Theory

The probability that an individual carries the putative deleterious gene depends on their family history of cancer, their age at diagnosis of cancer, if affected, or their age at last observation if unaffected. For a family of size m , where x_i denotes the i th individual's phenotype (affected or unaffected at a given age) the likelihood or probability of the observed phenotypes is given by:

$$\sum_{g_1} \dots \sum_{g_m} P(x_1, \dots, x_m | g_1, \dots, g_m) P(g_1, \dots, g_m)$$

where for each individual likelihoods are summed over all possible genotypes (g_i) (Ott, 1991). Thus the probability of an individual (m) being a carrier of a deleterious gene $P(D)$ under a dominant model is given by:

$$P(D) = \frac{\sum_{g_1} \dots \sum_{g_{m-1}} \sum_{\substack{g_m = dD \\ g_m = DD}} P(x_1, \dots, x_m | g_1, \dots, g_m) P(g_1, \dots, g_m)}{\sum_{g_1} \dots \sum_{g_m} P(x_1, \dots, x_m | g_1, \dots, g_m) P(g_1, \dots, g_m)}$$

i.e. in the numerator the genotype of individual m is fixed so that he or she is a carrier and the genotypes of other pedigree members vary given this assumption. Several computer programs are available to compute $P(D)$ (detailed below) but require input parameters corresponding to the probabilities of phenotype given genotype.

An individual's probability of being unaffected by cancer at age j given genotype g is given by the survival function:

$$S_g(j) = \exp \left(- \sum_{i=0}^{i=j} \lambda_{gi} \right) \tag{1}$$

where λ_{gi} is the genotype-specific incidence rate of cancer at age i and g refers to genotypes DD, Dd or dd . λ_{gi} is generally assumed to be constant over either a five or ten year age group. The complement of survivorship is the genotype-specific cumulative risk, given by:

$$F_g(j) = 1 - S_g(j) \tag{2}$$

An individual's probability of being affected by cancer at age j given genotype g is given by the genotype-specific density:

$$f_g(j) = \lambda_{gi} S_g(j) \tag{3}$$

For a dominant trait the values of S_{DD} and S_{Dd} , F_{DD} and F_{Dd} , and similarly f_{DD} and f_{Dd} are identical.

When no linkage information is available a consultand's risk between age i (current age) and j (future age) if he or she carries the deleterious gene (D) (defined here as the genetic risk) is given by:

$$(F_{DD}(j) - F_{DD}(i)) / (1 - F_{DD}(i)) \quad 4$$

and the consultand's risk if he or she is a non-carrier (defined here as the sporadic risk) is:

$$(F_{dd}(j) - F_{dd}(i)) / (1 - F_{dd}(i)) \quad 5$$

The consultand's overall risk is a weighted average of the genetic risk and the sporadic risk, given by:

$$P(D) \{(F_{DD}(j) - F_{DD}(i)) / (1 - F_{DD}(i))\} + (1 - P(D)) \{(F_{dd}(j) - F_{dd}(i)) / (1 - F_{dd}(i))\} \quad 6$$

For some common cancers a causative gene(s) has been identified, or chromosomal localisation of a gene(s) has been established. Provided the risk associated with being a carrier of an identified gene and the proportion of families which are linked have been estimated, risks can be computed in individual families which have a high probability of segregating the gene. The model of familial aggregation specified by segregation analysis will invariably reflect genetic heterogeneity and not mutations in a single gene. Consequently the consultand's overall risk will be composed of three elements: the risk associated with being a gene carrier under linkage; the risk associated with being a carrier under the segregation model but not a carrier at the linked locus (assuming the segregation model best predicts the effects of other unidentified genes); and the risk associated with not carrying a predisposition gene. The contribution of each of these to the overall risk will be a function of their respective probabilities. Hence the risk by age j for a consultand aged i who is an apparent carrier is given by:

$$\begin{aligned} & p \{(F_{LD}(j) - F_{LD}(i)) / (1 - F_{LD}(i))\} \\ & + (1 - p) P(D) \{(F_{DD}(j) - F_{DD}(i)) / (1 - F_{DD}(i))\} \\ & + (1 - p) (1 - P(D)) \{(F_{dd}(j) - F_{dd}(i)) / (1 - F_{dd}(i))\} \quad 7 \end{aligned}$$

$F_{LD}(i)$ and $F_{LD}(j)$ are the uncorrected cumulative risks of cancer in carriers at ages i and j respectively and p is the family's posterior probability of linkage defined by:

$$p = \alpha 10^{\text{Lod}} / (\alpha 10^{\text{Lod}} + (1 - \alpha)) \quad 8$$

Lod is the lod score for the family at the putative disease locus and α is the family's prior probability of linkage (i.e. the proportion of linked families of a similar type to the consultand's) (Terwilliger and Ott, 1994). Hence the probability of linkage takes into account the prior probability (assuming no linkage information) and the evidence in the family for linkage. $P(D)$ is assumed to approximate the consultand's risk of carrying a

predisposition gene given that the family is not segregating the gene for which linkage information is available. This may not be correct if the identified gene confers very different risks from those predicted by the segregation model. Similarly for a consultand aged i who does not carry the linked haplotype the risk by age j is given by:

$$\begin{aligned}
 & p \{ (F_{dd}(j) - F_{dd}(i)) / (1 - F_{dd}(i)) \} \\
 & + (1-p) P(D) \{ (F_{DD}(j) - F_{DD}(i)) / (1 - F_{DD}(i)) \} \\
 & + (1-p) (1-P(D)) \{ (F_{dd}(j) - F_{dd}(i)) / (1 - F_{dd}(i)) \}
 \end{aligned}
 \tag{9}$$

If a predisposing mutation has been identified within the family and the consultand can be tested, his or her risk simplifies to either the carrier risk if shown to have the mutation or approximately the sporadic risk if a non-carrier.

Where cancer at more than one site can be ascribed to the pleiotropic effects of a single gene, risks for the different sites can be computed by assuming site-specific incidence rates. It should be noted, however, that all risks are computed in the absence of death from other causes.

One caveat of using segregation data is that estimates of gene frequencies and hence risks, at older ages may be imprecise. This is because of a failure to incorporate genotype-specific mortality into any of the published analyses.

Practical implementation

For all but the simplest pedigree structures risk calculations will be complex and are therefore best carried out using computer programs. The computations required to determine the probability that the consultand is a gene carrier, given his or her age and family history under any segregation model, $P(D)$, can be undertaken using either the LINKAGE program (Lathrop and Lalouel, 1984; Lathrop *et al.*, 1984), or by modifying the USERM5 subroutine of the MENDEL program (Lange, 1988; Lange *et al.*, 1988). In both of these programs the penetrance probabilities assigned to liability classes are the genotype-specific densities for affecteds (equation 3). However, unaffecteds are treated differently, in LINKAGE the genotype-specific cumulative risks (equation 2) are used whereas the probabilities of being unaffected (equation 1) are the input parameters for the MENDEL program. The pedigree drawing program CYRILLIC allows input of pedigrees directly into both of these programs rather than entering the data into a file in the requisite format (Cyrillic for pedigree drawing, Chapman, 1995). This is simpler, more appropriate for use in clinical practice and probably less prone to errors in data entry than creating separate pedigree files. The other computations can be implemented using part of a genetic counselling program written in FORTRAN77 which we are currently developing for PC use. The part described in this article may be obtained through the anonymous f.t.p. server at venus.icr.ac.uk.

The breast cancer pedigree shown in Figure 1 provides an illustration of the aforementioned approach to estimate cancer risks. Table 1 shows the age-specific incidence rates of breast and ovarian cancer in non-carriers, in gene carriers under the CASH model and in BRCA1 gene carriers. The consultand's probability of carrying a breast/ovary gene, $P(D)$, under the CASH model based on her family history of cancer is 0.49 at age 25. (If unaffected at aged 65 her probability of being a carrier will fall to 0.25.) In the absence

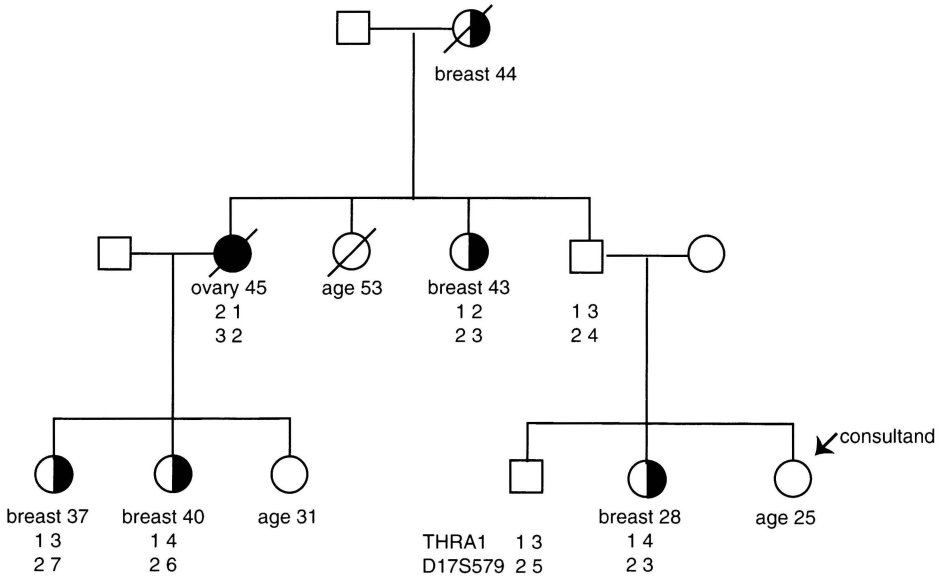


Figure 1: Pedigree of breast-ovarian cancer family. Chromosome 17q marker typings using THRA1 and D17S579 are shown.

of linkage information the consultand's estimated risks of breast and ovarian cancer between ages 25 and 50, are 19% and 2% respectively (using equation 6).

An estimated 81% of breast cancer families with one case of ovarian cancer and no male breast cancer are due to BRCA1 (Narod *et al.*, 1995). BRCA1 lies midway between markers THRA1 and D17S579 on chromosome 17. Typing shown in Figure 1 for the family gives a Lod score of 1.08 using the age-specific incidence rates for BRCA1 gene carriers and non-carriers in Table 1 and assuming 5 and 10 equally frequent alleles for THRA1 and D17S579 respectively. Using equation 8, the family's posterior probability of linkage is 0.98. The consultand's estimated risk of breast cancer between age 25 and 50, using equation 7, is 47% if she carries the linked haplotype and 1.7% if she does not. The corresponding risks for ovarian cancer are 15% and 0.3%.

DISCUSSION

Familial cancer risks can be estimated using the parameters obtained from segregation analyses alone or in conjunction with linkage data. The manipulations required to derive these estimates are readily carried out using a portable computer, so enabling risk estimates to be made available in clinical practice.

ACKNOWLEDGEMENTS

We would like to thank Cyril Chapman for advice. This work was supported by the Cancer Research Campaign.

Table 1: Age-specific breast and ovarian cancer incidence rates in non-carriers and carriers under the CASH model and in BRCA1 gene carriers.

Age group and disease state	Incidence rate / year		
	<u>Non-carriers</u>	<u>Carriers</u>	
	dd*	CASH model (dD or DD) [§]	BRCA1 (LD) [†]
Breast cancer < 30	0.000034	0.00168	0.0036
Breast cancer 30-39	0.00034	0.01391	0.016
Breast cancer 40-49	0.0011	0.03154	0.048
Breast cancer 50-59	0.0015	0.03222	0.034
Breast cancer 60-69	0.0019	0.03284	0.021
Breast cancer 70-79	0.0021	0.03284	0.023
Ovarian cancer < 30	0.000016	0.00027	0.0012
Ovarian cancer 30-39	0.000049	0.00083	0.0035
Ovarian cancer 40-49	0.00018	0.00306	0.013
Ovarian cancer 50-59	0.00036	0.00605	0.019
Ovarian cancer 60-69	0.00043	0.00734	0.019
Ovarian cancer 70-79	0.00046	0.00786	0.012

* Incidence rates for England and Wales 1978-1982 (Muir *et al.*, 1987).

§ Breast cancer incidence rates derived from Claus *et al.* (Risch and Thompson, 1991) but with modified rates for the age group 70-79 where rates were assumed to be the same as in the preceding age-group. Incidence rates for ovarian cancer were computed by assuming a constant relative risk with age and an estimated cumulative risk of 10% by age 60 (Claus *et al.*, 1993).

† Described in Narod *et al.*, 1995.

REFERENCES

- Chapman, C. (1995). *Cyrillic for pedigree drawing*. Cherwell Scientific Publishing Ltd., Oxford.
- Claus, E.B., Schildkraut, J.M., Thompson, W.D., Risch, N.J. (1993). Analysis of the genetic relationship between breast and ovarian cancer. *Am. J. Hum. Genet.*, **53**, A787.
- E.B., Risch, N.J., Thompson, W.D. (1991). Genetic analysis of breast cancer in the Cancer and Steroid Hormone Study. *Am. J. Hum. Genet.*, **48**, 232-242.
- Lange, K. (1988). *Documentation for Mendel version 2.2. programs for pedigree analysis*. Department of Biomathematics. University of California. Los Angeles.
- Lange, K., Weeks, D., Boehnke, M. (1988). Programs for pedigree analysis MENDEL, FISHER and dGENE. *Genet. Epidemiol.*, **5**, 471-472.
- Lathrop, G.M., Lalouel, J.M. (1984). Easy calculations of lod scores and genetic risks on small computers. *Am. J. Hum. Genet.*, **36**, 460-465.
- Lathrop, G.M., Lalouel, J.M., Julier, C., Ott, J. (1984). Strategies for multilocus linkage analysis in humans. *Proc. Natl. Acad. Sci. USA*, **81**, 3443-3446.
- Muir, C., Waterhouse, J., Mach, T., Powell, J., Whelan, S. (Eds). *Cancer Incidence in Five Continents*. Volume V, 1987. International Agency for Research on Cancer.

- Narod, S., Ford, D., Devilee, P., *et al.* (1995). An evaluation of genetic heterogeneity in 145 breast-ovarian cancer families. *Am. J. Hum. Genet.*, **56**, 254–264.
- Ott, J. (1991). *Analysis of Human Genetic Linkage*, 2nd ed. Johns Hopkins Press, Baltimore, Maryland.
- Terwilliger, J., Ott, J. (1994). *Handbook of Human Linkage Analysis*. Johns Hopkins University Press, Baltimore, Maryland.
- Weiss, K.M. (1993). *Genetic Variation and Human Disease. Principles and evolutionary approaches*. Cambridge University Press, Cambridge, Chapter 5, 67–90.



Hindawi
Submit your manuscripts at
<http://www.hindawi.com>

