Estimating penetrance from family data using a retrospective likelihood when ascertainment depends on genotype and age of onset.

Jérôme Carayol, Catherine Bonaïti-Pellié

▶ **To cite this version:**

Jérôme Carayol, Catherine Bonaïti-Pellié. Estimating penetrance from family data using a retrospective likelihood when ascertainment depends on genotype and age of onset.: Estimating penetrance from family data. Genetic Epidemiology, 2004, 27 (2), pp.109-17. 10.1002/gepi.20007 . inserm-00359199

HAL Id: inserm-00359199

https://inserm.hal.science/inserm-00359199

Submitted on 6 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimating penetrance from family data using a retrospective likelihood when ascertainment depends on genotype and age of onset

Jérôme Carayol and Catherine Bonaïti-Pellié

INSERM U535, Villejuif, France

Correspondance to:

Dr. Catherine Bonaïti-Pellié, INSERM U535, Bâtiment Leriche porte 18, BP 1000, 94817 Villejuif Cedex, France. Tel 33(0)1 45 59 53 49. Fax 33 (0)1 45 59 53 31 Email bonaiti@vjf.inserm.fr

ABSTRACT

In diseases caused by deleterious gene mutations, knowledge of age-specific cumulative risks is necessary for medical management of mutation carriers. When pedigrees are ascertained through several affected persons, ascertainment bias can be corrected by using a retrospective likelihood. This likelihood is a function of the genotypes of pedigree members given their phenotypes and provides unbiased estimates of penetrance without modeling the selection process, provided that selection is independent of genotypes. However, since mutation testing is offered only to relatives of mutation carriers, the genotypes of family members are available only in mutated families and selection does depend on genotype. In the present study, we quantified the bias due to selection on genotype using simulations. We found that this bias depended on the true penetrance value: the lower the penetrance, the higher the bias (risk by age 80 estimated to be 46% for a true penetrance value of 20%). When age of onset is added to the selection criteria, as usually done, we showed that the bias was even higher. We modified the conditioning in the retrospective likelihood - what we called "genotype restricted likelihood" (GRL). Using simulations, we showed that this method provided unbiased parameter estimates under all the selection designs considered.

Key words: Ascertainment, retrospective likelihood, genotype restricted likelihood (GRL), penetrance, pedigree analysis.

INTRODUCTION

There are now numerous examples of gene mutations predisposing to common diseases such as BRCA1 and BRCA2 in breast-ovarian cancer and mismatch repair (MMR) genes (MSH2, MLH1, MSH6, PMS1, PMS2) in colorectal cancer. Knowledge of age-specific cumulative risks of cancer (penetrance) for individuals who have inherited a deleterious mutation is necessary for medical management of mutation carriers. This estimation is difficult in diseases where only a minority of cases is due to such rare mutations. A good criterion for the presence of a mutation in a given individual is the existence of several affected relatives in his (her) family. For instance, the so-called "Amsterdam criteria" [Vasen et al., 1991; 1999] are widely used in the definition of hereditary nonpolyposis colorectal cancer (HNPCC) and include three close relatives with an HNPCC-associated cancer (colon, rectum, endometrium, small bowel, ureter or renal pelvis). These criteria have been shown to be highly predictive of a mutation of one of the MMR genes [Liu et al., 1996: Park et al., 2002]. The detection of a mutation in such families allows the identification of unaffected carriers who may thus undergo intensive surveillance, which considerably improves the prognosis of the disease. Some authors have used data on families ascertained on such criteria to estimate penetrance [Aarnio et al., 1995; Vasen et al., 1996; Voskuil et al., 1997; Lin et al., 1998; Aarnio et al., 1999]. However, ascertainment of multiple-case families was not taken into account in their analyses and Carayol et al. [2002] showed that this bias resulted in a large overestimation of risks.

Ascertainment-adjusted likelihood approaches allow unbiased estimation of penetrance on selected samples. When the ascertainment scheme is easy to model, a "prospective" likelihood [Hsu et al., 2000; Kraft and Thomas, 2000], which is a function of phenotypes (i.e.

disease status) given genotypes and ascertainment, may be used. However, when selection criteria are complex, this likelihood is difficult to compute. In this case, another approach, called "retrospective" likelihood [Hsu et al., 2000; Kraft and Thomas, 2000] may be used. The retrospective likelihood is a function of genotypes conditional on phenotypes. This method, related to the "MOD score" in linkage analysis [Clerget-Darpoux et al., 1986], can be applied whatever the selection criteria as the ascertainment process is not modeled in the likelihood. However, the retrospective likelihood corrects for ascertainment only if ascertainment does not depend on genotypes [Siegmund et al., 1999].

To estimate penetrance using a retrospective likelihood, families in which a mutation has been found in the frame of genetic counseling may be used. Practically, one of the affected family members (called index) is tested and if he (she) carries the mutation, the family members, affected or not, are offered genetic testing. Therefore, the assumption that selection is independent of genotype is violated and the retrospective likelihood does not adequately correct for ascertainment. Moreover, to increase the probability that the index case is not a sporadic case, an age criterion is often included in the selection criteria of families. For instance, in the Amsterdam criteria, one of the affected individuals must be affected before 50, and in the breast-ovarian cancer, the age criterion of 40 years is often proposed.

The aim of the present paper is first to quantify the bias on the penetrance estimates by using a retrospective likelihood when families are selected on carrier status, according to whether selection criteria include or not an age criterion for the index case, and then to propose a new likelihood which provides unbiased penetrance estimates under these selection designs.

METHODS

RETROSPECTIVE LIKELIHOOD

A retrospective likelihood is a function of observed genotypes, *Gen*, given observed phenotypes, *Phen*, ascertainment, *Asc*, and can be written *P(Gen/Phen,Asc)*. When ascertainment only depends on phenotypes, this likelihood reduces to P(*Gen*/*Phen*):

$$P(Gen/Phen,Asc) = \frac{P(Asc/Gen, Phen)\ P(Gen/Phen)}{P(Asc/Phen)} = P(Gen/Phen) \qquad (1)$$

Thus, the likelihood implicitly corrects for ascertainment without modelling the ascertainment process.

Let g and G denote the genotypes of non-carriers and carriers of the mutated allele respectively, Ω the set of all possible genotypic configurations for a family f, Gen_w the vector of genotypes in genotypic configuration w, $P(Gen_w)$ the corresponding probability and $P(Phen/Gen_w)$ the probability of the vector of observed phenotypes given the vector of genotypes in configuration w. The contribution of a given family f with s relatives can be written as :

$$L_f = P(Gen/Phen) = \frac{P(Phen_1,...,Phen_s/Gen_1,...,Gen_s)\ P(Gen_1,...,Gen_s)}{\sum_{w\in\Omega} P(Phen_1,...,Phen_s/Gen_{1,w},...,Gen_{s,w})\ P(Gen_{1,w},...,Gen_{s,w})} \qquad (2)$$

where the numerator of this likelihood is the joint probability of genotypes and phenotypes, and the denominator is the probability of all phenotypes of family f.

Under the assumption that phenotypes of relatives are independently distributed conditionally to their genotypes and that all genotypes are known:

$$L_f = P(Gen/Phen) = \frac{\prod_k P(Phen_k/Gen_k)\, P(Gen_1,\ldots,Gen_s)}{\sum_{w\in\Omega}\prod_k P(Phen_k/Gen_{k,w})\, P(Gen_{1,w},\ldots,Gen_{s,w})} \quad (3)$$

where $P(Phen_k/Gen_k)$ is the probability of phenotype ($Phen_k$) of individual k given his (her) genotype (Gen_k = G or g), and $P(Gen_1,\ldots,Gen_s)$, the joint probability of genotypes of the family. Given parents' genotypes, offspring genotypes are independent of each other, so

$$L_f = P(Gen/Phen) = \frac{\prod_k P(Phen_k/Gen_k)\prod_j P(Gen_j)\prod_{\{l,m,n\}} P(Gen_l/Gen_m, Gen_n)}{\sum_{w\in\Omega}\prod_k P(Phen_k/Gen_k)\prod_j P(Gen_{j,w})\prod_{\{l,m,n\}} P(Gen_{l,w}/Gen_{m,w}, Gen_{n,w})} \quad (4)$$

where the product on j is taken over all founders and the product on $\{l,m,n\}$ is taken over all parent-offspring triplets. Note that the denominator and the numerator are similar except that, in the denominator, the sum is taken over all possible genotypes.

If some individuals in family f are not tested, (4) can be written as:

$$L_f = P(Gen/Phen) = \frac{\sum_{z\in\Gamma}\prod_k P(Phen_k/Gen_{k,z})\prod_j P(Gen_{j,z})\prod_{\{l,m,n\}} P(Gen_{l,z}/Gen_{m,z}, Gen_{n,z})}{\sum_{w\in\Omega}\prod_k P(Phen_k/Gen_k)\prod_j P(Gen_{j,w})\prod_{\{l,m,n\}} P(Gen_{l,w}/Gen_{m,w}, Gen_{n,w})} \quad (5)$$

where Γ corresponds to the set of genotypic configurations compatible with genotypes of tested individuals.

For an individual i, $P(Gen_i)$ is expressed as a function of the frequency of the mutated allele in the general population assuming Hardy-Weinberg equilibrium for a founder (parents' status unknown). Otherwise, this probability depends on parental genotypes assuming Mendelian transmission.

Let $F_{Gen_i}(t)$ be the penetrance function at age t. If individual i is unaffected at age t_i, the contribution of i to the likelihood is:

$$P(Phen_i/Gen_i) = 1 - F_{Gen_i}(t_i)$$

that is the probability that individual i be still unaffected at age t_i (survival probability).

If individual i is affected at age t_i, the contribution of i to the likelihood is:

$$P(Phen_i/Gen_i) = F_{Gen_i}(t_i+1) - F_{Gen_i}(t_i)$$

that is the probability of being affected at age t_i included in the one year interval $[t_i ; t_i+1[$.

For the age-dependent penetrance function, one may choose a Weibull model with parameters λ (scale parameter) and a (shape parameter), because of its flexibility to adjust to observed data :

$$F_{Gen_i}(t) = 1-\exp[-(\lambda_{Gen_i} t)^{a_{Gen_i}}] \qquad (6)$$

QUANTIFICATION OF BIAS

The bias in the penetrance estimate due to genotype selection was quantified using simulations. For this purpose, samples of three-generation families in which a mutation segregates were simulated using various penetrance values in carrier individuals. The simulated pedigrees had a fixed structure and consisted of a couple of ancestors with two offspring and their spouses, having each four offspring. A carrier genotype was randomly assigned to one of the two ancestors. For the other family members, we simulated the genotype according to Mendel's laws for subjects whose parents were in the pedigree, and according to the population genotype frequency for spouses. The frequency of the mutated allele in the

population was arbitrarily fixed to 0.001 and we assumed the absence of *de novo* mutations. All genotypes were assumed to be known, and phenotypes were simulated according to the age dependent penetrance function using the Weibull model indicated above.

For sake of simplicity, whatever the carrier status, we fixed $a_G = a_g = 3$. For non-carriers ($Gen=g$), the parameter λ_g was set to a value corresponding to a cumulative risk of 0.02 by age 80. For carriers ($Gen=G$), we considered two different values for the parameter λ_G, the first one corresponding to a cumulative risk of 0.2 by age 80 (called "low true penetrance") and the second one to a cumulative risk of 0.5 by age 80 (called "high true penetrance"). We did not consider any gender difference in risks and all individuals were considered to be alive at the time of the study.

The families were selected if at least two relatives were affected. Two different designs were considered according to the existence or not of an age criterion. The age criterion was fixed to 40 years.

Let us call:

- design I: at least two affected relatives, one of whom is a carrier of the mutated allele.
- design II: at least two affected relatives, one of whom was affected before 40 years and is a carrier of the mutated allele.

Sample size was fixed to 10,000 families after selection to minimize sample fluctuations.

Parameters of the penetrance function were estimated by maximizing likelihood (4) using the Weibull model (6). A program has been written which uses the maximization

procedure GEMINI as a subroutine [Lalouel, 1979] and provides maximum likelihood estimates of the two parameters, λ_G and a_G, for carriers. We assumed that the penetrance was known for non-carriers and λ_g and a_g were fixed to the same values as in the simulation process of families.

CORRECTION OF THE LIKELIHOOD FOR SELECTION ON GENOTYPES

When family selection depends on genotypes, the simplification in (1) is not possible since *P(Asc/Gen,Phen)* is not equal to *P(Asc/Phen)*.

$$P(Gen/Phen,Asc) = \frac{P(Asc/Gen,Phen)\ P(Gen/Phen)}{P(Asc/Phen)} = \frac{P(Asc/Phen,Gen)\,P(Gen,Phen)}{P(Asc,Phen)} \qquad (7)$$

with

$$P(Asc,\ Phen) = \sum_{w \in \Omega} P(Asc/Phen, Gen_w)\ P(Phen,\ Gen_w) \qquad (8)$$

where Ω is composed of Ω_C, the set of genotypic configurations compatible with the selection criteria, and Ω_{NC}, the set of genotypic configurations not compatible with the selection criteria. Thus, $P(Asc/Phen,Gen_w) = 0$ if the family in genotypic configuration w does not fulfil the selection criteria. For a given family f with observed genotypic vector (*Gen*) and phenotypic vector (*Phen*), we can write:

$$P(Gen/Phen,\ Asc) = \frac{P(Asc/Gen,Phen)\ P(Phen/Gen)\ P(Gen)}{\sum_{w \in \Omega_C} P(Asc/Gen_w,\ Phen)\ P(Phen/Gen_w)\ P(Gen_w)} \qquad (9)$$

As demonstrated in the appendix, the probability of ascertainment given observed phenotypes and genotypes (*P(Asc/Gen,Phen)*) for a given family is identical whatever the genotypic

configuration. Thus,

$$P(Gen/Phen, Asc) = \frac{P(Phen/Gen)\,P(Gen)}{\sum_{w \in \Omega_C} P(Phen/\,Gen_w)\,P(Gen_w)} \qquad (10)$$

The likelihood can be developed as previously

$$P(Gen/Phen, Asc) = \frac{\prod_k P(Phen_k/Gen_k) \prod_j P(Gen_j) \prod_{\{l,m,n\}} P(Gen_l/Gen_m, Gen_n)}{\sum_{w \in \Omega_C} \prod_k P(Phen_k/Gen_k) \prod_j P(Gen_{j,w}) \prod_{\{l,m,n\}} P(Gen_{l,w}/Gen_{m,w}, Gen_{n,w})} \qquad (11)$$

The likelihood has the same form as the retrospective likelihood (4). Only the summation in the denominator is different. Practically, the denominator is computed as if all genotypes were unknown, using the Elston and Stewart algorithm [1971], but, when a given genotypic configuration is not compatible with the selection criteria, the family participation in this specific genotypic configuration is discarded from the denominator.

Subsequently, this likelihood (11) will be called “genotype restricted likelihood” (GRL) as it is a retrospective likelihood with conditioning on genotypic configurations restricted to genotypes compatible with selection criteria.

In order to validate the method, we applied the GRL on samples of simulated families selected under designs I and II, respectively. For samples of families selected under design I, we estimated the penetrance function using the GRL where Ω_C included the set of genotypic configurations with at least one affected family member being a carrier of the mutation (likelihood restricted on genotypes only referred as "GRL 1"). Under selection design II, Ω_C included the set of genotypic configurations with at least one affected family member being a carrier of the mutation and affected before 40 years (likelihood restricted on genotypes and age

of onset referred as "GRL 2").

CONFIDENCE INTERVALS

Let $\theta_{Gen,est} = (\lambda_{Gen,est}, a_{Gen,est})$ be the maximum likelihood estimates (MLEs) of θ_{Gen} and $F_{Gen,est}(t)$, the penetrance estimate obtained by replacing θ_{Gen} by $\theta_{Gen,est}$. To obtain a confidence interval of the penetrance function for carrier ($Gen = G$), asymptotic variance and covariance of the estimated parameters $\theta_{G,est}$ were obtained using GEMINI from the inverse of the Fisher information matrix I evaluated at the MLE, $I^{-1}(\theta_{G,est})$. Since $\theta_{G,est}$ is a MLE, $\theta_{G,est}$ converges to a normal distribution with mean θ_G and variance-covariance matrix $I^{-1}(\theta_{G,est})$. An approximation to the variance of the estimated penetrance was obtained by applying the Delta method:

$$\mathrm{Var}(F_{G,est}(t)) = D_{est}^{t}.I^{-1}(\theta_{G,est}).D_{est}$$

where D is the vector of partial derivatives of F with respect to each parameter:

$$D = (\delta F/\delta\lambda, \delta F/\delta a)^t$$

and D_{est} is calculated replacing θ_G by $\theta_{G,est}$ in D. A 95% confidence interval was obtained as in Brookmeyer and Crowley [1982]:

$$\{\ \theta_G : (F_{G,est}(t) - F_G(t))^2 \ \leq\ c\ .\ \mathrm{Var}\ (F_{G,est}(t))\ \}$$

where c is such as $\Pr(\chi_1^2 > c) = 0.05$.

RESULTS

Figures 1 and 2 show the penetrance estimates obtained by using a retrospective likelihood when ascertainment depends on genotypic status, in case of low and high true penetrance respectively. For low true penetrance (figure 1), when families were selected under design I, the estimated risks were at least twice the actual risks at all ages. Different results were obtained when risks were estimated from families selected under design II that included an age criterion for the index case. Compared to the results obtained under design I, the bias was higher, except in older age classes where estimates were quite similar. For high true penetrance (figure 2), the same trend was observed but the bias was smaller than for low true penetrance whatever the design. For still higher penetrance values, the bias was found to be negligible (data not shown). Additional analyses showed that the younger the age criterion, the greater the bias at intermediate ages (data not shown).

When the GRL 1 was used to estimate penetrance on families selected under design I, there was no more bias whatever the true penetrance value. Under design II, a first analysis was conducted using the GRL 1, i.e. using the same Ω_C as in design I. As shown in figure 3, for low true penetrance, the overestimation was smaller than when using the retrospective likelihood. For high true penetrance, the same observation was made until 70 years. After this age, the penetrance was slightly underestimated. When Ω_C was adequately defined as the set of genotypic configurations where at least one carrier individual was affected before age 40 (GRL 2), penetrance estimates were unbiased whatever the actual risks.

We studied the robustness of the GRL method to misspecification of the mutated allele frequency (q). When specifying a smaller value (q=0.0001), or a greater value (q=0.01) than the true one (q=0.001) in the likelihood, estimates remained quite stable under both designs and for both true penetrance values.

DISCUSSION

In the first part of this study, we showed that a retrospective likelihood did not adequately correct for ascertainment when only families with at least one carrier individual were used for estimating penetrance. This is consistent with the results of Siegmund et al. [1999] who showed that the MOD score provided biased penetrance estimates in carriers when the ascertainment was based on a LOD score criterion (i. e. ascertainment dependent on genotype).

In addition, we showed that introducing an age criterion in the selection criteria, which is usually done to increase the probability of detecting a mutation in families, worsened the bias due to the selection on genotype, in particular at young ages. We could check that introducing an age criterion had no effect when there was no selection on genotypes.

The GRL method that we propose in the present study to estimate penetrance from families selected through affected relatives and in which mutations have been detected, appears to be free of any bias. The GRL can be applied to any diseases where mutations have been identified as being responsible for Mendelian subentities. This method can be used whatever the selection criteria since the ascertainment process is not modeled in the likelihood. Thus, this method is adapted for estimating penetrance on samples of families selected using complex criteria involving many affected individuals, the probability of which is almost impossible to compute. This is generally the case in cancer syndromes such as hereditary nonpolyposis colorectal cancer associated to mutations in mismatch repair genes hMLH1 and hMSH2 or breast cancer associated to BRCA1 and BRCA2 mutations.

In our simulation study, we did not take mortality into account. Mortality is expected to decrease the overall information provided by a family as a smaller number of subjects would

be available, and in particular in older age, and thus to reduce the precision of estimates without introducing a bias. Similarly, a reduction in precision of estimates would be expected in case of unknown genotypes. In the simulations, we assumed that all genotypes were known. However, the GRL allows using observed phenotypes of individuals even if they are untyped. As the retrospective likelihood, the GRL is based on modeling genotypes given phenotypes. If only few relatives are typed, some problems of efficiency may be expected. We are currently investigating the properties of the GRLs in the case of unknown genotypes.

In this simulation study, we used sample of 10,000 families to avoid problems of sample fluctuations. Nevertheless, as seen in figures 1 to 3, we obtained rather large confidence intervals when using the retrospective or the GRL. In most studies, samples are much smaller, and it is clear that the GRL is subject to problems of efficiency on samples of realistic size. These results agree with those of Kraft and Thomas [2000] who showed that using the retrospective likelihood to correct for selection bias was at the expense of a loss of efficiency compared to other likelihoods.

The Weibull function, chosen to model the penetrance function, is widely used in parametric survival analysis because of its ability to adjust to observed data. The main advantage of this function (and more generally of parametric functions) is that only a small number of parameters are estimated to evaluate the penetrance function. Another way to estimate penetrance is to use linearly increasing risks within age classes, as in ARCAD [Le Bihan et al., 1995]. The assumption of linearity of risk can be justified if many small classes are used, but sometimes at the expense of a loss of efficiency in penetrance estimate for some age classes because of a lack of informative individuals. Furthermore, many classes imply a great number of parameters to estimate, with a possibility of stability problems. To avoid these problems, using few large age classes allows an increase of informativity within each class and

a reduction in the number of estimated parameters, with a possible lack of precision in penetrance estimate.

Regarding specification of mutated allele frequency q, we checked that the GRL was robust to a misspecification of this parameter. There was only a slight overestimation with high true penetrance when we introduced a value of q higher than the true value.

A retrospective likelihood, based on lod scores, was used for estimating breast and ovarian cancer risks associated with BRCA1 and BRCA2 mutations from multiple-case families collected through the Breast Cancer Linkage Consortium [Easton et al., 1993; Easton et al., 1995; Ford et al., 1998]. The penetrance estimates were generally higher in studies based on these multiple-case families than in those based on unselected series [Antoniou et al., 2003]. In order to explain such a difference, the existence of modifying factors, either genetic or environmental, that themselves run in families have been hypothesized [Easton, 1997] and several groups are now investigating this hypothesis. An alternative explanation is that studies on multiple-case families were biased by using a retrospective likelihood without taking into account the selection on mutation status. However, we have shown in this paper that this bias was negligible when penetrance was high and it is unlikely that the estimates obtained in these studies were affected by such a bias.

Finally, for all the Mendelian sub-entities of common diseases such as breast-ovarian cancer syndrome and HNPCC, cancer risks are always high and the bias evidenced in our study should not affect the estimates. This is no longer true when penetrance is low, and future studies on low penetrance mutations should absolutely correct for selection on genotypes and, when relevant, on age of onset by using GRLs such as the ones proposed in the present study.

APPENDIX

Asc, the ascertainment event, can be decomposed in two events: *Fsc*, the event that the family fulfilled the selection criteria, and *Rec*, the recruitment event. Thus,

$$P(Asc/Phen,Gen) = P(Fsc,Rec/Phen,Gen) = P(Rec/Fsc,Phen,gen)\ P(Fsc/Phen,Gen)$$

If a family fulfils the selection criteria, the information contained in *Fsc* is a subset of the information contained in vectors *Phen* and *Gen*, hence:

$$P(Fsc/Phen,Gen)=1$$

and

$$P(Rec/Fsc,Phen,Gen) = P(Rec/Phen,Gen)$$

$P(Rec/Phen,Gen)$ corresponds to the probability that family *f*, that fulfills the selection criteria, be recruited. Since recruitment occurs through affected individuals, this probability depends on the vector of observed phenotypes, *Phen*. Nevertheless, it is independent of the family genotypes distribution (*Gen*). Therefore:

$$P(Rec/Phen,Gen) = P(Rec/Phen)$$

Hence, the probability of ascertainment given observed phenotypes and genotypes ($P(Asc/Gen,Phen)$) for a given family is identical whatever the genotypic configuration.

ACKNOWLEDGEMENTS

We greatly acknowledge the support of the Fondation de France for Jérôme Carayol's fellowship.

REFERENCES

Antoniou A, Pharoah PDP, Narod S, Risch HA, Eyfjord JE, Hopper JL, Loman N, Olsson H, Johannsson O, Borg Å, Pasini B, Radice P, Manoukian S, Eccles DM, Tang N, Olah E, Anton-Culver H, Warner E, Lubinski J, Gronwald J, Gorski B, Tulinius H, Thorlacius S, Eerola H, Nevanlinna H, Syrjäkoski K, Kallioniemi O-P, Thompson D, Evans C, Peto J, Lalloo F, Evans DG, Easton DF. 2003. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet 72:1117-1130.

Aarnio M, Mecklin J-P, Aaltonen LA, Nyström-Lahti M, Järvinen HJ. 1995. Life-time risk of different cancers in hereditary non-polyposis colorectal cancer (HNPCC) syndrome. Int J Cancer 64:430-433.

Aarnio M, Sankila R, Pukkala E, Salovaara R, Aaltonen LA, de la Chapelle A, Peltomäki P, Mecklin J-P, Järvinen HJ. 1999. Cancer risk in mutation carriers of DNA-mismatch-repair genes. Int J Cancer 81:214-218.

Brookmeyer R, Crowley J. 1982. A confidence interval for the median survival time. Biometrics 38:29-41.

Carayol J, Khlat M, Maccario J, Bonaïti-Pellié C. 2002. Hereditary non-polyposis colorectal cancer: currents risks of colorectal cancer largely overestimated. J Med Genet 39:335-339.

Clerget-Darpoux F, Bonaïti-Pellié C, Hochez J. 1986. Effects of misspecifying genetic parameters in lod score analysis. Biometrics 42:393-399.

Easton DF, Bishop DT, Ford D, Crockford GP, and the Breast Cancer Linkage Consortium. 1993. Genetic linkage analysis in familial breast and ovarian cancer: results from 214

families. Am J Hum Genet 52:678-701.

Easton DF, Ford D, Bishop DT and the Breast Cancer Linkage Consortium. 1995. Breast and ovarian cancer incidence in BRCA1-mutation carriers. Am J Hum Genet 56:265-271.

Easton D. 1997. Breast cancer genes - What are the real risks ? Nat Genet 16:210-211.

Elston RC, Stewart J. 1971. A general model for the genetic analysis of pedigree data. Human Heredity 21:523-542.

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 BAJ, Gayther SA, Birch JM, Lindblom A, Stoppa-Lyonnet D, Bignon Y, Borg A, Hamann U, Haites N, Scott RJ, Maugard CM, Vasen H, Seitz S, Cannon-Albright LA, Schofield A, Zelada-Hedman M, and the Breast Cancer Linkage Consortium. 1998. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. Am J Hum Genet 62:676-689.

Hsu L, Ping Zhao L, Aragaki C. 2000. A note on a conditional-likelihood approach for family-based association studies of candidates genes. Hum Hered 50:194-200.

Kraft P, Thomas DC. 2000. Bias and efficiency in family-based gene-characterization studies: Conditional, Prospective, Retrospective, and joint likelihoods. Am J Hum Genet 66:1119-1131.

Lalouel JM. 1979. A computer program for optimization of general non linear functions. Technical report no 14. Salt Lake City: University of Utah, Department of Medical Biophysics and Computing.

Le Bihan C, Moutou C, Brugières L, Feunteun J, Bonaïti-Pellié C. 1995. ARCAD: a method for estimating age-dependent disease risk associated with mutation carrier status from family data. Genet Epidemiol 12:13-25.

Lin KM, Shashidharan M, Thorson AG, Ternent CA, Blatchford GJ, Christensen MA, Watson P, Lemon SJ, Franklin B, Karr B, Lynch J, Lynch HT. 1998. Cumulative incidence of colorectal and extracolonic cancers in MLH1 and MSH2 mutation carriers of hereditary nonpolyposis colorectal cancer. J Gastrointest Surg 2:67-71.

Liu B, Parsons R, Papadopoulos N, Nicolaides NC, Lynch HT, Watson P, Jass JR, Dunlop M, Wyllie A, Peltomäki P, de la Chapelle A, Hamilton SR, Vogelstein B, Kinzler KW. 1996. Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients. Nat Med 2:169-174.

Park JG, Vasen HFA, Park YJ, Park KJ, Peltomaki P, Ponz De Leon M, Rodriguez-Bigas MA, Lubinski J, Beck NE, Bisgaard M-L, Miyaki M, Wijnen JT, Baba S, Lindblom A, Madlensky L, Lynch HT. 2002. Suspected HNPCC and Amsterdam criteria II: evaluation of mutation detection rate, an international collaborative study. Int J Colorectal Dis 17:109-114.

Siegmund KD, Gauderman WJ, Thomas DC. 1999. Gene characterization using high-risk families: a sensitivity analysis of the MOD score approach. Am J Hum Genet Suppl 65:A398.

Struewing JP, Hartge P, Wacholder S, Baker SM, Berlin M, McAdams M, Timmerman MM, Brody LC, Tucker MA. 1997. The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. N Engl J Med 336:1401-1408.

Vasen HFA, Mecklin J-P, Meera Khan P, Lynch HT. 1991. The International Collaborative

Group on hereditary non-polyposis colorectal cancer (ICG-HNPCC). Dis Colon Rectum 34:424-425.

Vasen HFA, Wijnen JT, Menko FH, Kleibeuker JH, Taal BG, Griffioen G, Nagengast FM, Meijers-Heijboer EH, Bertario L, Varesco L, Bisgaard M-L, Mohr J, Fodde R, Meera Khan P. 1996. Cancer risk in families with hereditary nonpolyposis colorectal cancer diagnosed by mutation analysis. Gastroenterology 110:1020-1027.

Vasen HF, Watson P, Mecklin JP, Lynch HT, and the ICG-HNPCC. 1999. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the international collaborative group on HNPCC. Gastroenterology 116:1453-1456.

Voskuil DW, Vasen HFA, Kampman E, Van't Veer P, and the national collaborative group on HNPCC. 1997. Colorectal cancer risk in HNPCC families: development during lifetime and in successive generations. Int J Cancer 72:205-209.

LEGENDS FOR FIGURES

Fig. 1. Penetrance function in case of low true penetrance (thick solid line) estimated using a retrospective likelihood from families selected under design I (fine solid line) and under design II (dotted line), and 95% confidence interval.

Fig. 2. Penetrance function in case of high true penetrance (thick solid line) estimated using a retrospective likelihood from families selected under design I (fine solid line) and under design II (dotted line), and 95% confidence interval.

Fig. 3. Penetrance function estimated using the GRL 1 from families selected under design II (fine lines), and 95% confidence interval, depending on true penetrance (thick lines): dotted line for high true penetrance and solid line for low true penetrance.