

## Research Article

# Age-Dependent Cancer Risk Is Not Different in between *MSH2* and *MLH1* Mutation Carriers

Sylviane Olschwang,<sup>1,2</sup> Kai Yu,<sup>3</sup> Christine Lasset,<sup>4</sup>  
Stéphanie Baert-Desurmont,<sup>5</sup> Marie-Pierre Buisine,<sup>6</sup> Qing Wang,<sup>4</sup>  
Pierre Hutter,<sup>7</sup> Etienne Rouleau,<sup>8</sup> Olivier Caron,<sup>9</sup>  
Violaine Bourdon,<sup>2</sup> and Gilles Thomas<sup>10</sup>

<sup>1</sup>Institut National de la Santé et de la Recherche Médicale (INSERM), Unité 891, Centre de Recherches en Cancérologie de Marseille, 13009 Marseille, France

<sup>2</sup>Department of Oncogenetics, Institut Paoli-Calmettes, 13009 Marseille, France

<sup>3</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute (NCI), National Institutes of Health (NIH), Gaithersburg, MD 20877, USA

<sup>4</sup>Oncogenetics Department, Centre Léon Bérard, 69003 Lyon, France

<sup>5</sup>Medical Genetics Department, Hôpital Charles-Nicolle, 76000 Rouen, France

<sup>6</sup>Molecular Biology Laboratory, Hôpital Claude Huriez, 59000 Lille, France

<sup>7</sup>Institut Central des Hôpitaux Valaisans, CH-1951 Sion, Switzerland

<sup>8</sup>Oncogenetics Department, Centre René Huguenin, 92210 Saint-Cloud, France

<sup>9</sup>Genetics Department, Hospices Civils, 67000 Strasbourg, France

<sup>10</sup>Fondation Synergie Lyon Cancer, 69008 Lyon, France

Correspondence should be addressed to Sylviane Olschwang, sylviane.olschwang@inserm.fr

Received 22 July 2008; Revised 17 November 2008; Accepted 9 January 2009

Recommended by Thomas L. Vaughan

Lynch syndrome is mostly characterized by early-onset colorectal and endometrial adenocarcinomas. Over 90% of the causal mutations occur in two mismatch repair genes, *MSH2* and *MLH1*. The aim of this study was to evaluate the age-dependent cancer risk in *MSH2* or *MLH1* mutation carriers from data of DNA diagnostic laboratories. To avoid overestimation, evaluation was based on the age-dependent proportion of mutation carriers in asymptomatic first-degree relatives of identified mutation carriers. Data from 859 such eligible relatives were collected from 8 centers; 387 were found to have inherited the mutation from their relatives. Age-dependent risks were calculated either using a nonparametric approach for four discrete age groups or assuming a modified Weibull distribution for the dependence of risk on age. Cancer risk was estimated starting at 28 (25–32 0.68 confidence interval) and to reach near 0.70 at 70 years. The risks were very similar for *MSH2* and *MLH1* mutation carriers. Although not statistically significant, the risk in males appeared to precede that for females by ten years. This difference needs to be investigated on a larger dataset. If confirmed, this would indicate that the onset of the colonoscopic surveillance may be different in male and female mutation carriers.

Copyright © 2009 Sylviane Olschwang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. Introduction

Many genetic disorders have been found to exhibit a simple Mendelian inheritance pattern, and advances in the knowledge about their genetic basis have led to the expansion of DNA testing both for diagnosis and prediction of disease susceptibilities. In the case of mutations associated with an increased risk of common cancers, one parameter of major practical importance is the age-dependent cancer risk,

defined as the risk for a mutation carrier of developing a tumor prior to a given age. Indeed, a precise knowledge of this parameter is instrumental in the counseling of individuals who are identified as carriers by genetic testing and who are faced with different options for cancer prevention or early detection.

A number of methodologies have been developed to estimate penetrance and the lifetime risk or recurrence

risk of cancer prone individuals. Often, the first evaluation of cancer risk is performed from symptomatic individuals identified in large pedigrees used in linkage studies [1]. It has been shown that such design leads to a severe overestimation. Alternative evaluation methods, which tend to reduce such biases, include population-based studies and/or prospective follow-up of unaffected mutation carriers. These approaches, however, may be expensive, time consuming, and may require long follow-up in order to provide sufficient information. HNPCC, also known as the Lynch syndrome, is an autosomal dominant condition caused by mutation in one or several genes involved in DNA mismatch repair (MMR) [2]. Mutation carriers have been shown to be at high risk to develop colorectal and endometrial adenocarcinomas. In addition, significantly increased risks have been reported for cancers of small bowel, upper urological tract, stomach, ovary, and biliary tract [3]. Although at least four MMR genes (*MSH2*, *MLH1*, *MSH6*, and *PMS2*) have been implicated in Lynch syndrome, more than 90% of the causative mutations have been identified in two of them, *MSH2* and *MLH1*. It has been estimated that the prevalence of mutations in these two MMR genes in the general population of European origin is between 1 of 500 and 1 of 1000 [4]. The prevalence in colorectal cancer patients is 2.7% [4]. In studies where ascertainment of Lynch families was not corrected, the estimated lifetime risk of colorectal cancer ranges from 0.68 to 0.82.

A precise knowledge of the age-dependent risk of cancer for individuals with deleterious *MSH2* and *MLH1* mutations is helpful in the identification and clinical management of families at high risk of colorectal and endometrial cancers. However, it has been recognized that evaluation of the cancer risk of HNPCC individuals performed from symptomatic patients referred to a cancer family clinic leads to overestimation [1, 5, 6]. We had previously briefly delineated an evaluation method which may be less sensitive to recruitment bias. It is based on the age-dependent proportion of mutation carriers observed in asymptomatic offspring of mutation carriers which when applied to 267 individuals led to an evaluation of the age-dependent risks of first cancer to be approximately 0.43 at age 38 and 0.62 at age 51 in mutation carriers [7]. The recent development of cancer family clinics offers the potential to generate a large amount of data, thus providing the opportunity of improving evaluations of cancer risk. Here, we more explicitly present the method and provide an example of its application by studying data from a total of 859 asymptomatic offspring of mutation carriers, distributed over an extended range of age, that have been collected through the contribution of hospital laboratories which perform genetic testing of *MSH2* and *MLH1* genes in France and Switzerland. We also show that the number of observations has to be substantially increased in order to provide precise estimates.

## 2. Patients and Methods

**2.1. Patients.** A retrospective questionnaire was sent to eight genetic units which offer germline analysis of *MSH2* and

*MLH1* genes under a Health Ministry agreement in France and Switzerland. This questionnaire asked, for each genetic test performed on asymptomatic offspring of mutation carriers, the following information: disease causing germline mutation identified in the proband using the international mutation nomenclature (<http://www.hgvs.org/mutnomen/>), birth date, sex, and age at genetic diagnosis. The questionnaire was fulfilled by the biologist having validated the predictive tests. No follow-up data of the corresponding at-risk relatives was required.

**2.2. Genetic Testing.** In all laboratories that provide data, the presence or absence of the disease causing mutation was assessed on genomic DNA extracted from two independent blood samples according to the French and Swiss rules for examination of the individual genetic characteristics. Depending on the mutation type, point mutation, or large genomic rearrangement found in the proband, either by DNA sequencing or by quantitative, fluorescent multiplex PCR [8] was performed.

**2.3. Risk Calculation.** The method is based on the determination of the age-dependent proportion of mutation carriers observed in asymptomatic first-degree relatives of mutation carriers at the time of the genetic test. No question about survival of mutation carrier is addressed in this work. The probabilities at birth for a first-degree relative of a mutation carrier to be either mutation carrier or nonmutation carrier are approximately equal and will be assumed equal in the rest of the study. For these two groups, the proportion of individuals that become symptomatic with age differs. Let  $\pi_{ng}(t)$  and  $\pi_{gc}(t)$  be the probability of a nonmutation carrier and of a mutation carrier to be affected by cancer before age  $t$ , respectively.  $\pi_{gc}(t)$  is also called cancer risk. The proportions  $r_{ng}$  and  $r_{gc}$  of asymptomatic individuals that can still be observed at age  $t$  are  $r_{ng} = 1 - \pi_{ng}(t)$  and  $r_{gc} = 1 - \pi_{gc}(t)$ . Therefore, in proportion of mutation carriers at age  $t$  in a group of asymptomatic first-degree relatives of mutation carriers;  $p(t) = r_{gc}/(r_{ng} + r_{gc})$ , it follows that

$$\pi_{gc}(t) = \frac{1 - p(t) - p(t)(1 - \pi_{ng}(t))}{1 - p(t)}. \quad (1)$$

It is, therefore, possible to evaluate the age-dependent increased risk of mutation carriers (and, therefore, cancer risk) from the age-dependent risk of nonmutation carriers and the age-dependent proportion of mutation carriers among asymptomatic first-degree relatives of mutation carriers. In the rest of this work, we will assume that  $\pi_{ng}(t)$  remains small so that we have the approximate relationship:

$$\text{Cancer risk} = \pi_{gc}(t) \cong \frac{1 - 2p(t)}{1 - p(t)}. \quad (2)$$

A nonparametric estimate of the cancer risk can be obtained as  $(1 - 2\hat{p}(t))/(1 - \hat{p}(t))$ , where  $\hat{p}(t)$  is an estimation of  $p(t)$  approximated by  $N_{as}/(N_{ng} + N_{as})$ , where  $N_{as}$  (and  $N_{ng}$ ) is the number of mutation carriers (and noncarriers) sampled at age between  $[t - \Delta t, t + \Delta t]$ . Thus,  $\pi_{gc}(t)$  is estimated as

TABLE 1: Distribution of mutation carrier and non carrier in 859 unaffected individuals with one 1st degree relative carrying a deleterious *MSH2* or *MLH1* mutation.

Age group	Mutation absent				Mutation present			
	Females		Males		Females		Males	
	<i>MLH1</i>	<i>MSH2</i>	<i>MLH1</i>	<i>MSH2</i>	<i>MLH1</i>	<i>MSH2</i>	<i>MLH1</i>	<i>MSH2</i>
<30	47	30	52	26	52	38	35	47
30 to 40	36	30	39	28	35	30	25	16
40 to 50	23	25	18	20	18	17	14	10
50 to 60	12	11	20	18	8	10	8	9
> 60	15	10	10	2	7	2	2	4

$(N_{ng} - N_{as})/N_{ng}$ , with variance  $N_{as}(N_{ng} + N_{as})/N_{ng}^3$  according to the  $\delta$  method [9].

A parametric estimate based on the logistic regression model can also be proposed. Based on the definition of  $\pi_{gc}(t)$  and  $p(t)$ , we have the following important relationship:

$$\log \frac{p(t)}{1-p(t)} = \log [1 - \pi_{gc}(t)]. \quad (3)$$

Assuming the following modified Weibull distribution for the cancer risk:

$$\pi_{gc}(t) = \begin{cases} 0, & \text{if } t \leq \tau, \\ 1 - \exp(-[a(t-\tau)]^b), & \text{if } t > \tau, \end{cases} \quad (4)$$

we observe that

$$\log \frac{p(t)}{1-p(t)} = \log [1 - \pi(t)] = -a[(t-\tau)^+]^b. \quad (5)$$

This equation suggests the familiar logistic regression model. We can consider the following simple model:

$$\log \frac{p(t)}{1-p(t)} = \beta(t-\tau)^+. \quad (6)$$

We can find the maximum likelihood estimate of  $(\beta, \tau)$  by maximizing the likelihood function based on the above probability model. Once we have an estimate for  $(\beta, \tau)$ , we can obtain the estimate for  $\pi_{gc}(t)$  for any given  $t$ . We call this estimate the parametric estimate. To obtain the confidence interval of the parametric estimate, we use the standard bootstrap method [10]. To compare the cancer risk between two groups, such as males and females, we used the likelihood ratio statistic that compares the likelihood assuming the same parametric model (i.e., the common  $(\beta, \tau)$  for both samples) with the likelihood obtained by allowing  $(\beta, \tau)$  to be varied between two samples. The statistical significance of the test can be evaluated through a permutation procedure by randomly shuffling the group ids (i.e., gender or gene name) among all subjects.

### 3. Results

Eight genetic units contributed information on a total of 859 asymptomatic offspring of mutation carriers: 581 from SO,

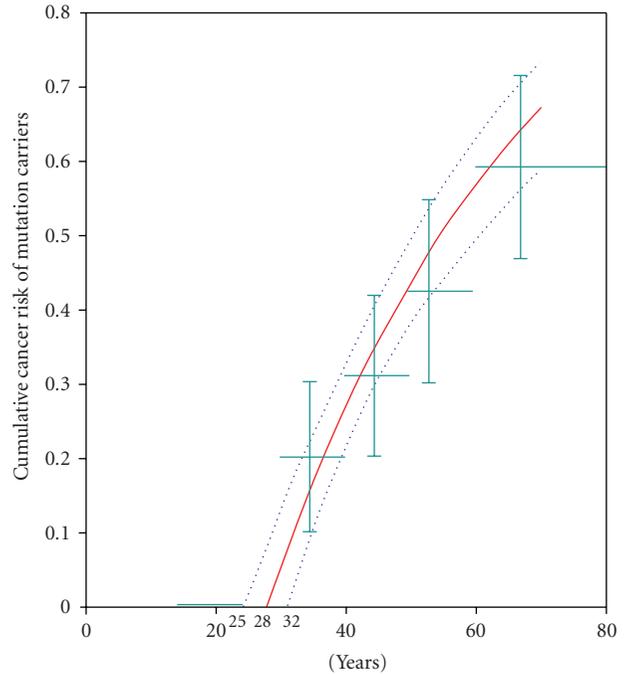


FIGURE 1: Cumulative risk of cancer for *MSH2* and *MLH1* mutation carriers. Individuals are not differentiated with respect to the mutated gene or gender. The five horizontal blue lines indicate the risk for each of the 5 groups as shown in Table 1. For the youngest group, the risk was evaluated to zero. For the other four groups, the vertical line is placed at the median age and indicates the 0.68 confidence interval for the evaluated risk. The red curve shows the risk evaluated assuming that it follows a modified Weibull distribution as a function of age. The two flanking dotted graphs indicate the 0.68 confidence interval of this evaluation. The age at onset of the risk is evaluated to 28 (0.68 confidence interval = 25–32 years).

116 from CL and QW, 58 from SBS, 44 from MPB, 21 from PH, 16 from ER, 14 from OC, and 9 from VB. They were 472 nonmutation carriers (233 males and 239 females, aged 18 to 89 years) and 387 mutation carriers. Mutation was located within the *MSH2* gene in 183 cases (86 males and 97 females, aged 18 to 73 years) and within the *MLH1* gene in 204 cases (84 males and 120 females, aged 18 to 74 years). Mutations were all predicted to result in protein truncation or were missense mutations classified as deleterious on the basis of

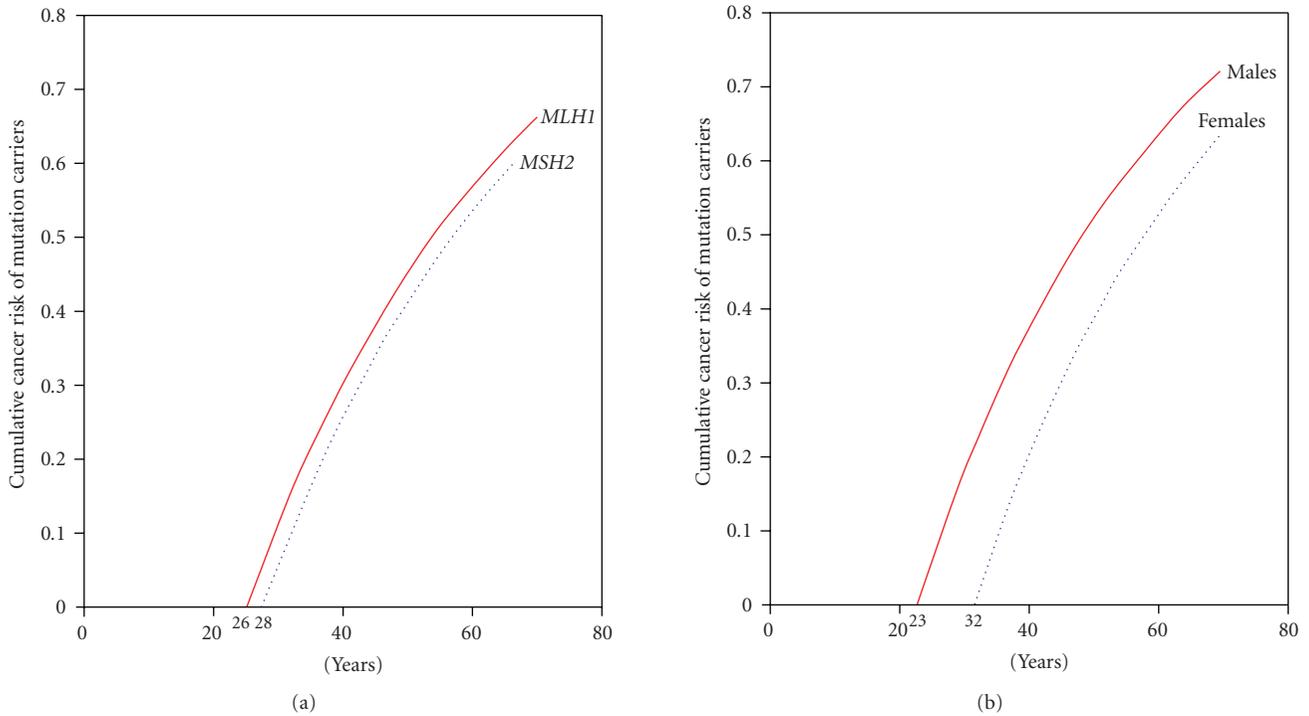


FIGURE 2: Cumulative risk of cancer for mutation carriers when individuals are differentiated with respect to the mutated gene (part (a)) or gender (part (b)). The risk estimation was only attempted under the assumption that the risk follows a modified Weibull distribution as a function of age.

functional tests in yeast, cosegregation analyses, and tumor cells studies. First-degree relatives of index cases carrying DNA variants of unknown significance were removed from this study. The 859 unaffected first-degree relatives of *MSH2* and *MLH1* mutation carrier were classified into five age groups (Table 1).

When both the gender and the nature of the mutated gene (whether *MSH2* or *MLH1*) are considered, the number of observations in each group is too small to enable a non-parametric evaluation of the cancer risk. Pooling together the four groups (*MHS2*, *MLH1*, males, and females), we attempted a nonparametric evaluation of the age-dependent cancer risk. It can be observed that as pooled age groups become older, the proportions of affected mutation carriers tend to decrease. This decrease is due to the removal from the study of the mutation carriers who become symptomatic. In the younger age group, the number of mutation carriers was larger than that of the nonmutation carriers, an observation that is likely due to the small number of observations and which suggests that cancer risk is very small in this age group. For the other groups, the proportion of mutation carriers was smaller than 0.5, and thus for these groups, a nonzero cancer risk could be estimated (Figure 1). For instance, for the age group between 50 and 60, the median age was 53, and the cancer risk was evaluated to 0.43. We note that the standard deviation of the present evaluation is large.

In an attempt to obtain a more precise evaluation, we performed a parametric estimate of the age-dependent

cancer risk assuming that cancer risk would be negligible before an age threshold called  $\tau$ , and starting from this age, cancer risk would increase according to a Weibull distribution with a parameter  $\beta$ . Weibull distributions are currently used in survival analyses and have been applied to parameterize age-dependent cancer risk [11]. Under this model, the maximum likelihood estimate of  $\tau$  is 28 years (68% confidence interval 25–32 years). After this age, cancer risk rises rapidly and reaches a value of 0.48 (68% confidence interval 0.42–0.54) at 53 years, and 0.67 (68% confidence interval 0.59–0.74) at 70 years. After this age, the probability of a nonmutation carrier to have developed cancer becomes substantial so that the model may need to be corrected according to (1). The number of observations of nonsymptomatic first-degree relatives older than 70 is small in our series, and no attempt was made to evaluate cancer risk after this age.

A similar method was applied separately on asymptomatic individuals with first-degree relatives carrying an *MLH1* or an *MHS2* mutation. The difference is risk for *MSH2* and *MLH1* mutation carriers appeared minimal (Figure 2(a)). When males and females were analyzed separately, the age-dependent risk for males appeared shifted by ten years as compared to females. Under the parametric model, the age at onset of the increased risk in males is 23 years (95% confidence interval 11–27) as it is 32 years (95% confidence interval 29–37) in females. However, a permutation test failed to demonstrate statistical significance ( $P = .15$ ).

#### 4. Discussion

There is a clear need to improve our estimation of the age-dependent cancer risk for many genetic diseases. This is especially important for those conditions that predispose to cancer as this knowledge may influence the definition of the best surveillance protocol. The development of presymptomatic DNA diagnostic tests offers an opportunity to improve this knowledge. However, we need to apply methods that are less prone to biases than those based on the age at the onset of symptomatic individuals referred to cancer family clinics [11, 12].

The evaluation method discussed in this paper requires a set of data that are collected in a two-stage process. In the first stage, symptomatic individuals are referred to a clinic and the deleterious mutations are identified. Importantly, the age at onset of the symptomatic individuals collected at this stage is not used to evaluate cancer risk as it is well known that such procedure may lead to major overestimation. In the second stage, asymptomatic first-degree relatives of individuals with an identified mutation are recruited and a test is conducted to determine their mutation status. Cancer risk is only evaluated from the age-dependent proportion of gene carriers in this group of asymptomatic first degree relatives. This method shares some of the potential biases that may be observed in population-based studies. The highly penetrant mutations are likely to contribute more than the low-penetrant mutations as the earlier are more readily detected than the former. Also mutations that lie in chromosomal regions that are investigated by routine DNA diagnostic techniques (e.g., mainly exonic point mutations or genomic large rearrangements) have been preferentially included in the study. Thus, the group of mutations that have been evaluated for cancer risk may not be representative of the mutation spectrum that is present in the population. However, for the group of mutations that have been identified, the method appears minimally biased. This lack of bias stems from the requirement that the individuals included in the study should be asymptomatic.

The present method requires a large number of observations. With the present dataset, variance of our estimation is large. It is barely informative when a nonparametric evaluation method is used (e.g., when individuals are pooled into 10-year age groups). The assumption of a modified Weibull distribution for cancer risk enables to decrease this variance at the cost of minimal hypotheses. Simulation studies indicate that the collection of a 4-fold increased the number of observations would decrease the confidence interval by a factor 2 (results not shown). With the development of presymptomatic DNA diagnostic tests, such number should be obtainable in the near future at little cost.

In the present work, we have applied the method to the evaluation of the age-dependent cancer risk of mutations in the *MSH2* and *MLH1* genes associated with Lynch syndrome. The resulting evaluation of the age-dependent cancer risk is consistent with those that have been previously published based on population-based studies [13]. It does confirm that the previous evaluation based on the age at onset of retrospectively included symptomatic individuals was

overestimated. It also indicates that as previously proposed, cancer risk of *MLH1* and *MSH2* mutations is similar [14]. An analysis distinguishing gender also suggests that the onset of the increased risk occurs ten years earlier in males than in females. This observation, if confirmed, suggests that the colonoscopy surveillance in males may have to be started earlier than in females, possibly leading to changes in the standard guidelines [15, 16]. Also, if it is confirmed that cancer risk is lower at all ages in females than in males, it would imply that the colorectal risk may be much smaller in females than in males since females are also at high risk of endometrial cancer. Similar observations have been recently published in the literature [12, 14].

Besides Lynch syndrome, it would be of interest to apply this approach to predisposing diseases for which the first manifestations are not present at birth and which may have irreversible deleterious consequences when the diagnosis is delayed until symptomatic. This includes not only the cancer predisposing conditions such as those associated to BRCA mutations [17, 18] but also possibly conditions associated with other degenerative processes (neurological or metabolic).

#### Acknowledgments

This work was supported by Institut National du Cancer and the French ministry of health. SO coordinates the MMR Genes Testing National Network.

#### References

- [1] H. F. A. Vasen, J. T. Wijnen, F. H. Menko, et al., "Cancer risk in families with hereditary nonpolyposis colorectal cancer diagnosed by mutation analysis," *Gastroenterology*, vol. 110, no. 4, pp. 1020–1027, 1996.
- [2] S. Syngal, E. A. Fox, C. Eng, R. D. Kolodner, and J. E. Garber, "Sensitivity and specificity of clinical criteria for hereditary non-polyposis colorectal cancer associated mutations in *MSH2* and *MLH1*," *Journal of Medical Genetics*, vol. 37, no. 9, pp. 641–645, 2000.
- [3] H. F. A. Vasen, P. Watson, J.-P. Mecklin, and H. T. Lynch, "New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative Group on HNPCC," *Gastroenterology*, vol. 116, no. 6, pp. 1453–1456, 1999.
- [4] R. Salovaara, A. Loukola, P. Kristo, et al., "Population-based molecular detection of hereditary nonpolyposis colorectal cancer," *Journal of Clinical Oncology*, vol. 18, no. 11, pp. 2193–2200, 2000.
- [5] P. Watson and H. T. Lynch, "Extracolonic cancer in hereditary nonpolyposis colorectal cancer," *Cancer*, vol. 71, no. 3, pp. 677–685, 1993.
- [6] M. Aarnio, R. Sankila, E. Pukkala, et al., "Cancer risk in mutation carriers of DNA-mismatch-repair genes," *International Journal of Cancer*, vol. 81, no. 2, pp. 214–218, 1999.
- [7] Y. Parc, C. Boisson, G. Thomas, and S. Olschwang, "Cancer risk in 348 French *MSH2* or *MLH1* gene carriers," *Journal of Medical Genetics*, vol. 40, no. 3, pp. 208–213, 2003.
- [8] F. Charbonnier, G. Raux, Q. Wang, et al., "Detection of exon deletions and duplications of the mismatch repair genes in hereditary nonpolyposis colorectal cancer families using

- multiplex polymerase chain reaction of short fluorescent fragments,” *Cancer Research*, vol. 60, no. 11, pp. 2760–2763, 2000.
- [9] J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley Series in Probability and Statistics, John Wiley & Sons, New York, NY, USA, 2002.
- [10] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Chapman and Hall, New York, NY, USA, 1993.
- [11] Y.-H. Choi, K. A. Kopciuk, and L. Briollais, “Estimating disease risk associated with mutated genes in family-based designs,” *Human Heredity*, vol. 66, no. 4, pp. 238–251, 2008.
- [12] F. Quehenberger, H. F. A. Vasen, and H. C. van Houtwelingen, “Risk of colorectal and endometrial cancer for carriers of mutations of the hMLH1 and hMSH2 gene: correction for ascertainment,” *Journal of Medical Genetics*, vol. 42, no. 6, pp. 491–496, 2005.
- [13] M. G. Dunlop, S. M. Farrington, A. D. Carothers, et al., “Cancer risk associated with germline DNA mismatch repair gene mutations,” *Human Molecular Genetics*, vol. 6, no. 1, pp. 105–110, 1997.
- [14] E. Barrow, W. Alduaij, L. Robinson, et al., “Colorectal cancer in HNPCC: cumulative lifetime incidence, survival and tumour distribution. A report of 121 families with proven mutations,” *Clinical Genetics*, vol. 74, no. 3, pp. 233–242, 2008.
- [15] N. M. Lindor, G. M. Petersen, D. W. Hadley, et al., “Recommendations for the care of individuals with an inherited predisposition to Lynch syndrome: a systematic review,” *Journal of the American Medical Association*, vol. 296, no. 12, pp. 1507–1517, 2006.
- [16] H. F. A. Vasen, G. Möslein, A. Alonso, et al., “Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer),” *Journal of Medical Genetics*, vol. 44, no. 6, pp. 353–362, 2007.
- [17] K. Offit, “BRCA mutation frequency and penetrance: new data, old debate,” *Journal of the National Cancer Institute*, vol. 98, no. 23, pp. 1675–1677, 2006.
- [18] R. Fu, E. L. Harris, M. Helfand, and H. D. Nelson, “Estimating risk of breast cancer in carriers of *BRCA1* and *BRCA2* mutations: a meta-analytic approach,” *Statistics in Medicine*, vol. 26, no. 8, pp. 1775–1787, 2007.



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

