

Strategies for screening for hereditary non-polyposis colorectal cancer

Anu Loukola, Albert de la Chapelle, Lauri A Aaltonen

Abstract

Germline mutations in DNA mismatch repair genes (*MLH1*, *MSH2*, *PMS1*, *PMS2*, and *MSH6*) predispose to hereditary non-polyposis colorectal cancer (HNPCC). In the absence of pathognomonic clinical features, diagnosis of HNPCC is often based on family history. Microsatellite instability (MSI) analysis has successfully been used for screening colorectal cancer patients for HNPCC. The aim of this study was to evaluate the feasibility of a recently introduced logistical model based on family history data in detecting HNPCC patients with germline mutations. A series of 509 kindreds with a proband with colorectal cancer was studied. MSI analysis and subsequent germline mutation analysis (*MLH1* and *MSH2*) in MSI positive patients had been performed previously. Of the 509 patients, 63 (12%) were MSI positive and 10 (2%) had a germline mutation in *MLH1* or *MSH2*. The power of the logistical model was tested to determine its value in predicting the probability of a germline mutation. The model proposed a high probability in three out of 10 mutation positive cases when data on cancer in first degree relatives were considered (typically three generation pedigrees, consisting, on average, of eight people). Using extended pedigrees and family cancer data in the 10 mutation positive kindreds (an average of 38 family members available), the model suggested high probabilities in seven out of 10 mutation positive cases. We conclude that for the model to predict germline mutation cases, extensive pedigrees and family history data are required. When screening colorectal cancer patients for HNPCC, a model using a combination of family information and MSI has optimal specificity and sensitivity.

(J Med Genet 1999;36:819-822)

Keywords: HNPCC; screening; MSI; colon cancer

Each year in the United States some 160 000 people are diagnosed with colorectal cancer (CRC) and it is the primary cause of death in some 60 000.¹ Approximately 15% of CRCs are familial. In the majority of these, the

underlying mechanism of predisposition is not known. However, two distinct heritable conditions account for a proportion of the familial cases. Familial adenomatous polyposis (FAP) is the underlying cause in less than 1% of all CRC,² while the hereditary non-polyposis colorectal cancer (HNPCC) syndrome accounts for 2 to 5%.³⁻⁵ In HNPCC, efficient early detection of tumours by repeated colonoscopy prevents cancer deaths.⁶ It is therefore of considerable interest to identify high risk subjects, that is, those who have germline mutations. In both FAP and HNPCC the lifetime risk of cancer is between 80 and 100%.^{2,7-9} In FAP, subjects at high cancer risk are often readily identifiable because their colonic mucosa displays numerous visible benign polyps. By contrast, in HNPCC, high risk subjects cannot be identified by clinical means. Deficient mismatch repair is the cause of HNPCC.¹⁰ Carriership for a germline mutation of one of the mismatch repair genes (*MLH1*, *MSH2*, *PMS1*, *PMS2*, or *MSH6*) constitutes a diagnosis of HNPCC and can be determined by analysing the relevant genes for mutations.

Given that all patients newly diagnosed with CRC need to be evaluated for HNPCC, how can this be done in a rational, efficient, and reasonably cost efficient way? Two recent papers addressed this issue.^{5,11} Aaltonen *et al*⁵ used the microsatellite instability (MSI) phenotype to screen 509 consecutive newly diagnosed colorectal tumours and found 63 that were positive. Mutation analyses of the two main mismatch repair genes (*MLH1* and *MSH2*) disclosed a germline mutation in 10 of the 63 MSI unstable patients. Thus, 10/509 patients (2%) were diagnosed as having HNPCC in this series. The authors emphasised the use of MSI as a primary molecular screening method.

Wijnen *et al*¹¹ analysed a previously collected series of 184 families in which HNPCC or an HNPCC-like condition conferred high risk of CRC. To predict the probability of finding a disease causing germline mutation in *MLH1* or *MSH2*, the authors proposed an algorithm in which the variables were mean age of CRC diagnosis in the family, fulfilment of certain family history criteria, and the presence of endometrial cancer in the family. Applying this

Department of
Medical Genetics,
Haartman Institute,
University of Helsinki,
PO Box 21, FIN-00014
Helsinki, Finland
A Loukola
L A Aaltonen

Comprehensive
Cancer Center, The
Ohio State University,
Columbus, Ohio, USA
A de la Chapelle

Correspondence to:
Dr Aaltonen.

Revised version received
27 May 1999
Accepted for publication
17 June 1999

algorithm (<http://www.nfdht.nl>), a probability of 20% or higher for a germline mutation was proposed to justify mutation analysis.

Can the two strategies be consolidated, as one might deduce from the comment by Lynch and Smyrk?⁴ There is indeed an urgent need to devise an acceptable strategy. If the logistical model proposed¹¹ could be used to define all or most high risk subjects among all CRC cases, one would not need a laboratory test to select those high risk subjects who would require mutation analysis. In this way large scale mutation screening might be both efficient and cost effective.

Materials and methods

PATIENTS AND TISSUE PREPARATION

A consecutive series of 509 fresh frozen colorectal adenocarcinomas were collected and prepared as previously described.⁵ The study was approved by the appropriate ethical committee and informed consent was obtained from each patient before any molecular analyses were carried out. The first degree relatives of the patients were identified through population registries, and information on the cancer status of each of the subjects was derived from the Finnish Cancer Registry or through death certificates. Including each proband, an average of eight first degree relatives were identified, typically in three generations. In mutation positive cases the pedigrees were extended further. Of the 509 probands eight had a first degree relative with endometrial cancer, 65 had a first degree relative with colorectal cancer, and an additional six probands had both endometrial and colorectal cancer in their first degree relatives.

ANALYSIS OF MICROSATELLITE INSTABILITY AND GERMLINE MUTATIONS

Analysis of microsatellite instability was performed with several mono-, di-, and tetranucleotide markers as previously described.⁵ MSI

positive patients were analysed for germline mutations in *MLH1* and *MSH2* by denaturing gradient gel electrophoresis or direct genomic sequencing.⁵

TESTING THE MATHEMATICAL MODEL

Mathematical analysis was performed according to Wijnen *et al.*¹¹ The probability of detecting *MLH1* or *MSH2* mutations in individual families was calculated as previously proposed, using the following equation: $p = e^L / (1 + e^L)$, where e is the exponential function and L is the log odds. The log odds can be calculated with the following formula: $L = 1.4 + (-0.1)V_1 + 1.7V_2 + 2.4V_3$, where V_1 is the mean age at diagnosis of colorectal cancer of all affected members of the family, V_2 equals 1 if at least one member of the family has endometrial cancer and equals 0 otherwise, and V_3 equals 1 if the family meets the Amsterdam criteria and equals 0 otherwise.

Because the Amsterdam criteria are often regarded as too restrictive for small families, an alternative formula was also proposed. The log odds for the alternative formula can be calculated with the following formula: $L = 1.4 + (-0.09)V_1 + 0.27V_2 + 0.75V_3$, where V_1 is the mean age at diagnosis of colorectal cancer of all affected members of the family, V_2 is the number of patients with colorectal cancer in the family, and V_3 is the number of patients with endometrial cancer in the family. A probability of 20% or higher was proposed to be sufficient to justify germline mutation analysis (*MLH1* and *MSH2*).¹¹

The data derived from the 509 kindreds were evaluated using both versions of the model. In the cases where a germline mutation was known to exist, extended pedigree data were also available and used during a second round of analyses, to determine the effect of this additional information on the outcome. The program available at <http://www.nfdht.nl> gives slightly different values from the formula

Table 1 Calculations using the algorithms of Wijnen *et al.* to determine the probabilities of finding a germline *MLH1* or *MSH2* mutation in 10 mutation carrier probands previously ascertained by Aaltonen *et al.*

Patient	Mean age of diagnosis of CRC in the family	Fulfilment of the Amsterdam criteria	Endometrial cancer in the family	No of patients with CRC	No of patients with endometrial cancer	No of family members with verified cancer status	p_1	p_2
<i>(A) Limited pedigrees (proband and the first degree relatives)</i>								
c35	49	No	No	2	0	6	3	8
c54	36	No	No	1	0	4	10	17
c64	53	No	No	2	0	5	2	6
c125	52	No	No	2	0	9	2	6
c145	42	Yes	Yes	3	1	12	78	31
c219	50	No	Yes	1	1	7	13	11
c340	61	No	Yes	3	2	7	5	14
c430	38	No	No	2	0	4	8	19
c558	57	No	Yes	4	2	8	7	24
c564	45	No	Yes	2	1	7	19	20
<i>(B) Extended pedigrees</i>								
c35	47	Yes	No	4	0	24	30	15
c54	35	No	No	1	0	13	11	19
c64	47	Yes	No	4	0	24	10	11
c125	48	Yes	No	4	0	14	28	14
c145	56	Yes	Yes	13	3	85	48	89
c219	52	No	Yes	7	2	49	11	53
c340	56	Yes	Yes	5	2	33	48	31
c430	44	Yes	Yes	9	1	33	75	65
c558	57	Yes	Yes	5	3	39	45	47
c564	45	Yes	Yes	12	4	77	73	97

p_1 = probability of finding a germline mutation in *MLH1* or *MSH2*.

p_2 = probability of finding a germline mutation in *MLH1* or *MSH2*; the alternative formula.

presented in the article by Wijnen *et al.*¹¹ As there was no program available at <http://www.nfdht.nl> for the alternative model, we calculated all probabilities using the formulae as shown in that article.

Results

The results are summarised in table 1. When the p_1 formula¹¹ was applied, only one out of the 10 HNPCC mutation carriers showed a probability of over 20% (in this case 78%). In the remaining nine cases the probabilities ranged between 2% and 19%. Using the alternative formula (p_2), only three patients showed probabilities of 20% or higher (20%, 24%, and 31%), while in the remaining seven patients the probabilities ranged between 6% and 19%. These findings were based on family histories comprising only first degree relatives, typically in three generations (parents, sibs, children). The mean number of first degree family members in whom cancer status could be verified was 8.1 (range 1 to 18) (table 1A).

To examine whether the availability of more extensive information increases the accuracy of prediction, we again applied the formula but with the inclusion of verified cancer data on a mean of 38.4 (range 13 to 85) family members (table 1B). In this case, the p_1 value exceeded 20% in seven probands, but was below 20% in three probands (10%, 11%, and 11%). Using the p_2 formula, four patients had probabilities lower than 20% (11%, 14%, 15%, and 19%).

Thus using the proposed algorithm¹¹ a 20% cut off limit failed to identify most (seven to nine) mutation positive subjects if data were available only on first degree family members, and failed to identify one third (three to four) mutation positive subjects even with extensive pedigree and cancer information.

We then asked the question whether applying the algorithm to all 509 patients in our series would have identified subjects at high risk whose tumours were not MSI positive, and therefore had not been subjected to mutation analyses.⁵ Three such subjects were identified. One of these (probability p_1 17% and p_2 32%) had FAP and one (probability 27/38%) had juvenile polyposis. The DNA of the third patient (probability 15/25%) was analysed for mutations in *MLH1* and *MSH2* by genomic sequencing of the promoter region and the coding exons, but was mutation negative.

Discussion

In a series of 509 colorectal cancer patients studied previously, all 10 newly diagnosed HNPCC probands had at least one of three characteristics: below 50 years of age (7/10), at least one first degree family member with colorectal or endometrial cancer (9/10), and a previous colorectal or endometrial cancer in the proband (3/10).⁵ Had these criteria been used to select patients for MSI analysis, 122 (24%) of the 509 tumours would have been studied for MSI, 20 would have been positive, and by mutational analysis, of these 20 subjects, the same 10 probands would have been found. Thus, applying the above selection criteria, the number of cases needing labora-

Table 2 The impact of different cut off levels on the number of probands to be selected for germline mutation analysis using the algorithm of Wijnen *et al.*¹¹ The total number of probands scrutinised was 509

Cut off level	Mutation positive probands exceeding the cut off level		Total number of probands exceeding the cut off level	
	p_1	p_2	p_1	p_2
20%	1/10	3/10	2	6
15%	2/10	5/10	4	8
10%	4/10	7/10	6	23
5%	6/10	10/10	21	52

tory investigations would be reduced but the sensitivity would remain high. Here, we evaluated the selection criteria proposed by Wijnen *et al.*¹¹ which are based on similar clinical and family history information.

The mathematical model¹¹ has not been tested previously in a prospectively collected series of colon cancer kindreds. Our experience confirms the easy use of the model, but suggests that in the setting of first degree family histories the cut off level of 20% probability for germline testing is too high. We calculated the effects of lowering the cut off level for germline mutation testing (table 2). The sensitivity increased dramatically only when the cut off level was lowered to 5%, reaching 100% when the alternative formula (p_2) was used. At the same time the number of patients requiring mutation testing also increased dramatically (table 2). However, in the studied series the specificity can be considerably improved by performing MSI analysis in those patients whose probability (p_2) exceeds 5% and selecting only MSI positive patients for mutation analyses. In this study, 52 out of 509 colon cancer patients exceeded 5% probability, 12 of them were MSI positive, and the 10 mutation positive subjects were among these.

In most circumstances probands can provide relatively reliable data on first degree family members and cancer status can be ascertained in most of these relatives. More extensive pedigree information is increasingly difficult to obtain, though inclusion of second degree relatives has been recommended.¹² This approach is complicated by the fact that the accuracy of cancer data obtained from the probands varies by site, being especially low in abdominal malignancies.¹² In the present study, even use of the most extensive pedigree information with verified cancer data was not sufficient to detect one third of the mutation positive cases.

What, then, is at present the optimal strategy to screen newly diagnosed CRC patients for HNPCC? The goal must be to maximise both sensitivity (as few false negative results as possible) and specificity (as few tests as possible) of any method applied on a large scale. The data presented here lend themselves to the following concrete proposal. As extensive pedigree information and verified data on cancer in distant family members cannot always be obtained, we propose routinely to consider first degree relatives only (this does not preclude the use of more extensive family data whenever such data are available). Moreover, we propose using MSI as a prescreening test in every case, as this

appears to carry only a 5-15% false negative rate.^{4,13} Based on this, two methods are possible. (A) Perform MSI testing on tumours from probands with at least one of the three following criteria: (1) age under 50 years, (2) at least one first degree relative of the proband with CRC or endometrial cancer, (3) a previous CRC or endometrial cancer in the proband. Search for mutations in those patients whose tumours are MSI positive. Based on Aaltonen *et al*⁷ this implies performing MSI analysis in 24% of all patients and mutation analysis in 4%. (B) The alternative scenario is to use the algorithm (p₂) of Wijnen *et al*¹¹ and perform MSI analysis on tumours of all patients whose risk is 5% or greater and mutational analysis in those whose tumours are MSI positive. This implies doing MSI in only 10% and mutational analyses in 2.4% of all patients, but the approach relies on extensive and accurate pedigree and cancer information.

Scenario A maximises the sensitivity, especially whenever the pedigree information is weak. Scenario B maximises specificity, but carries the risk of a loss of sensitivity. There is clearly a need for further prospective studies to assess the relative merits of different strategies. It may well be that different strategies will turn out to be optimal in different settings. Variables that may be of importance include the overall incidence of CRC in the population under study, the proportion of HNPCC, the existence of easily detectable widespread founder mutations, the availability and coverage of cancer registries, the attitudes of the public, and the coverage of insurance. Moreover, the compelling need to restrict the number of patients

being offered mutation analysis because of its high cost may change when easier methods become available.

We thank Siv Lindroos and Sinikka Lindh for assistance. This study was supported by grants from the Finnish Cancer Society, the Academy of Finland, the Sigrid Juselius Foundation, Duodecim, and the National Institutes of Health (CA67941 and CA16058), and by contract (BMH4-CT96-0772) with the European Commission.

- 1 Parker SL, Tong T, Bolden S, Wingo PA. Cancer statistics, 1996. *CA Cancer J Clin* 1996;46:5-27.
- 2 Phillips RKS, Spigelman AD, Thompson JPS, eds. *Familial adenomatous polyposis and other polyposis syndromes*. London: Edward Arnold, 1994.
- 3 Mecklin JP, Järvinen HJ, Hakiluoto A, *et al*. Frequency of hereditary nonpolyposis colorectal cancer. A prospective multicenter study in Finland. *Dis Colon Rectum* 1995;38:588-93.
- 4 Lynch HT, Smyrk TC. Identifying hereditary nonpolyposis colorectal cancer. *N Engl J Med* 1998;338:1537-8.
- 5 Aaltonen LA, Salovaara R, Kristo P, *et al*. Incidence of hereditary nonpolyposis cancer and the feasibility of molecular screening for the disease. *N Engl J Med* 1998;338:1481-7.
- 6 Järvinen H, Mecklin JP, Sistonen P. Screening reduces colorectal cancer rate in families with hereditary nonpolyposis colorectal cancer. *Gastroenterology* 1995;108:1405-11.
- 7 Dunlop MG, Farrington SM, Carothers AD, *et al*. Cancer risk associated with germline DNA mismatch repair gene mutations. *Hum Mol Genet* 1997;6:105-10.
- 8 Vasen HF, van Ballegooijen M, Buskens E, *et al*. A cost-effectiveness analysis of colorectal screening of hereditary nonpolyposis colorectal carcinoma gene carriers. *Cancer* 1998;2:1632-7.
- 9 Aarnio M, Sankila R, Pukkala E, *et al*. Cancer risk in mutation carriers of DNA mismatch repair genes. *Int J Cancer* 1999;81:214-18.
- 10 Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. *Cell* 1996;18:159-70.
- 11 Wijnen JT, Vasen HFA, Khan PM, *et al*. Clinical findings with implications for genetic testing in families with clustering of colorectal cancer. *N Engl J Med* 1998;339:511-18.
- 12 Douglas FS, O'Dair LC, Robinson M, Evans DGR, Lynch SA. The accuracy of diagnosis as reported in families with cancer: a retrospective study. *J Med Genet* 1999;36:309-12.
- 13 Aaltonen LA, Peltomäki P, Mecklin JP, *et al*. Replication errors in benign and malignant tumors from hereditary nonpolyposis colorectal cancer patients. *Cancer Res* 1994;54:1645-8.