

Research

Open Access

Penetrance of colorectal cancer among *MLH1/MSH2* carriers participating in the colorectal cancer familial registry in Ontario

Yun-Hee Choi^{1,2}, Michelle Cotterchio^{3,4}, Gail McKeown-Eyssen^{5,6}, Monga Neerav³, Bharati Bapat^{7,8}, Kevin Boyd³, Steven Gallinger^{7,9,10}, John McLaughlin^{1,3,5}, Melyssa Aronson¹¹ and Laurent Briollais*^{1,5}

Address: ¹Prosserman Centre for Health Research, Samuel Lunenfeld Research Institute, Toronto, Canada, ²Department of Epidemiology and Biostatistics, University of Western Ontario, London, Canada, ³Division of Preventive Oncology, Cancer Care Ontario, Toronto, Canada, ⁴Division of Population Studies and Surveillance, Cancer Care Ontario, Toronto, Canada, ⁵Dalla Lana School of Public Health, University of Toronto, Toronto, Canada, ⁶Department of Nutritional Sciences, University of Toronto, Toronto, Canada, ⁷Fred A. Litwin Centre for Cancer Genetic, Samuel Lunenfeld Research Institute, Toronto, Canada, ⁸Laboratory of Medicine and Pathobiology, Mount Sinai Hospital, Toronto, Canada, ⁹Department of Surgery, University of Toronto, Toronto, Canada, ¹⁰Ontario Familial Colon Cancer Registry, Ontario Cancer Genetics Network, Toronto, Canada and ¹¹Dr. Zane Cohen Digestive Disease Clinical Research Centre, Mount Sinai Hospital, Toronto, Canada

Email: Yun-Hee Choi - Yun-Hee.Choi@schulich.uwo.ca; Michelle Cotterchio - Michelle.Cotterchio@cancercare.on.ca; Gail McKeown-Eyssen - gail.eyssen@utoronto.ca; Monga Neerav - Neerav.Monga@cancercare.on.ca; Bharati Bapat - Bapat@lunenfeld.ca; Kevin Boyd - Kevin.Boyd@cancercare.on.ca; Steven Gallinger - sgallinger@rogers.com; John McLaughlin - John.Mclaughlin@cancercare.on.ca; Melyssa Aronson - maronson@mtsina.on.ca; Laurent Briollais* - laurent@lunenfeld.ca

* Corresponding author

Published: 23 August 2009

Received: 13 May 2009

Hereditary Cancer in Clinical Practice 2009, **7**:14 doi:10.1186/1897-4287-7-14

Accepted: 23 August 2009

This article is available from: <http://www.hccjournal.com/content/7/1/14>

© 2009 Choi et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Several DNA mismatch repair (MMR) genes, responsible for the majority of Lynch Syndrome cancers, have been identified, predominantly *MLH1* and *MSH2*, but the risk associated with these mutations is still not well established. The aim of this study is to provide population-based estimates of the risks of colorectal cancer (CRC) by gender and mutation type from the Ontario population.

Methods: We analyzed 32 families segregating MMR mutations selected from the Ontario Familial Colorectal Cancer Registry and including 199 first-degree and 421 second-degree relatives. The cumulative risks were estimated using a modified segregation-based approach, which allows correction for the ascertainment of the Lynch Syndrome families and permits account to be taken for missing genotype information.

Results: The risks of developing CRC by age 70 were 60% and 47% among men and women carriers of any MMR mutation, respectively. Among *MLH1* mutation carriers, males had significantly higher risks than females at all ages (67% vs. 35% by age 70, p -value = 0.02), while the risks were similar in *MSH2* carriers (about 54%). The relative risk associated with *MLH1* was almost constant with age (hazard ratio (HR) varied between 5.5-5.1 over age 30-70), while the HR for *MSH2* decreased with age (from 13.1 at age 30 to 5.4 at age 70).

Conclusion: This study provides a unique population-based study of CRC risks among *MSH2/MLH1* mutation carriers in a Canadian population and can help to better define and understand the patterns of risks among members of Lynch Syndrome families.

Introduction

Lynch Syndrome, also referred to as Hereditary non-polyposis colorectal cancer (HNPCC) is an autosomal dominant condition which predisposes carriers to both colorectal and extra-colonic cancers [1]. Several DNA mismatch repair (MMR) genes responsible for the majority of Lynch Syndrome cancers have been identified, predominantly *MLH1* and *MSH2* [2]. An understanding of risk associated with MMR mutations is important to assist in decisions about prophylactic surgery, annual screening and chemoprevention and to help allay psychological distress related to the uncertainty of colorectal cancer (CRC) predisposition among Lynch Syndrome family members [1].

Despite this importance, only a small number of studies provided penetrance estimates for these gene mutations. In addition, most of them studied populations from Finland, Scotland, the Netherlands and Nebraska [3-6] and it is unlikely that they are generalizable to the Canadian population as genetic mutations, lifestyle and environmental factors may differ. Moreover, most previous studies focused on families with a very strong history of cancers associated with the Lynch Syndrome and the derived penetrance estimates are only valid for families with comparable family history. Our study provides population-based estimates of penetrance by age, sex and mutated gene for CRC cases identified through the Ontario Familial Colorectal Cancer Registry (OFCCR) [7] and selected to be representative of families likely to harbour MMR mutations. We also used an ascertainment-corrected segregation approach to correct for the specific ascertainment and missing mutation information. Therefore, our analyses reduce the chance for biases and provide penetrance estimates that are applicable to a larger population of CRC cases harbouring MMR mutations.

Materials and methods

Study population

The OFCCR is one of six international registries established by the National Cancer Institute (NCI) as a resource for the study of CRC [7]. The OFCCR used the population-based Ontario Cancer Registry (OCR) to identify incident CRC cases (proband), aged 20-74, diagnosed July 1997 to July 2000. With physician consent, colorectal cases were contacted by a letter describing the study and requesting completion of a family history questionnaire (FHQ) which ascertained birth, death and disease history for first- and second-degree relatives. This information was used by genetic counselors to produce family pedigrees and to classify families as either 1) Amsterdam 1, 2) Familial Risk which included early-onset CRC and/or extra-colonic cancer in proband or relative, or family history of multiple CRC and/or extra-colonic cancers, 3) Sporadic/low risk with no family history of CRC [7].

Proband from Amsterdam 1 and Familial Risk families were interviewed by counselors to clarify information regarding all first-degree relatives, complete the pedigree to include all second-degree relatives, and ascertain full cancer histories. A randomized sample of 25% of the sporadic/low risk cases was also studied but because the prevalence of MMR germline mutations is very low in the general population, around 1:3139 [8], this group was not further considered in the penetrance study. Indeed, only one individual in this group was found mutation carrier after tumor analysis and germline testing.

Whenever possible, probands' tumours were first screened for microsatellite instability (MSI) and immunohistochemistry (IHC). Germline mutation analysis of *MSH2*, *MLH1* and/or *MSH6* genes was initiated in probands based on MSI-high and/or IHC deficient tumours results. Germline screening of a specific MMR gene was guided by IHC deficiency for a specific MMR protein. Germline mutations were also assessed among probands with MSI-low or microsatellite stable (MSS) tumours if their tumours exhibited MMR protein deficiency when assessed by IHC [9]. Proband were also screened for germline mutations in *MLH1* and *MSH2* when the tumours were MSI-high and IHC intact, or where tumours were not available for probands in Amsterdam 1 families. Testing for MMR genes involved sequence analysis (*MSH2*, *MLH1*, *MSH6*) and Multiplex Ligation-dependent Probe Amplification (MLPA) (*MSH2*, *MLH1*) of genomic DNA obtained from blood samples [9].

For each proband found to carry an MMR mutation, all first- and second-degree relatives on the "at-risk side" of the proband's pedigree were considered to be eligible for the present study. Blood samples from these relatives were obtained in two stages. First, the OFCCR had requested blood samples from all probands' first-degree relatives, together with all second-degree relatives who had had cancer or who were a first-degree relative of a family member with cancer, and who resided in North America, most European countries, or the Caribbean. Second, to ensure as many kin as possible were ascertained for the present study, we requested blood from all probands' second-degree relatives who had not previously been contacted by the OFCCR. Inclusion of these additional second-degree relatives was important to avoid over representation of relatives with a cancer diagnosis and resultant upwardly biased risk estimates.

Mutations in the MMR genes were ascertained from genomic DNA for all participants from whom blood samples could be obtained. For deceased relatives, obligate gene mutation carriers were defined as persons with a descendent with a MMR gene mutation. Therefore, the final retrospective study cohort included all living *MSH2*,

MLH1 and *MSH6* mutation carriers, and all deceased obligate gene mutation carriers. There were 446 probands eligible for our study (i.e., with blood sample available, with consent to contact their relatives, satisfying the Amsterdam 1 or other familial risk criteria, see above). Of these, 358 were tested for IHC (80%), 346 for MSI (77%), and 379 for either IHC or MSI (85%). Our final cohort included 32 probands found to carry a MMR mutation and their relatives. There were 27 probands from Amsterdam 1 families and 5 from other familial risk families. The relatives consisted of 199 first-degree and 421 second-degree relatives, from whom 56 and 41, respectively, were CRC cases (see Table 1). The number of blood samples available was 71 and 56, respectively, in first- and second-degree relatives from whom 38 and 22 were tested mutation positive (see Table 1). Survival analyses with the classical Kaplan-Meier estimator [10] were carried out on the sample of individuals with blood available. The modified segregation-based approach (see methods section) was able to use the full cohort of individuals (i.e. with or without blood available) but age at diagnosis or at examination was required for the analysis. This was available for 506 individuals of the 622 recruited.

The Ethic approval was obtained from Mount Sinai Hospital Research Ethics Board for this research project.

Verification of cancer diagnoses in relatives carrying a MMR gene mutation

The pathology and date of diagnosis of all subject-reported CRC in a first- or second-degree relative were verified where possible using pathology reports obtained from the OCR, other cancer registries outside of Ontario, hospital discharge data, death certificates, and reports from Regional Cancer Centres and Princess Margaret Hospital, the main cancer hospital in Toronto. If only a death certificate was available, it was reviewed. We attempted to verify cancer diagnoses reported among relatives living in other countries by requesting a pathology report from the relevant cancer registry or hospital. All probands had their colorectal cancer diagnosis confirmed by pathology reports (100%), as required for study eligibility. Of the 622 first- and second-degree relatives identified in the 32 mutation positive families, 61% of CRCs identified were confirmed by medical records and 38% of relatives identified as deceased were confirmed by death records found in the mortality database.

Identification of probands/families carrying a MMR gene mutation

Microsatellite instability (MSI) analysis

Written informed consent was obtained from 99.6% of OFCCR participants for the collection of tumour tissue for use in cancer research. Colorectal tumour blocks were collected from all eligible patients. Tumour DNA was extracted from paraffin embedded matched normal and

tumour tissue specimens and tested for MSI status, using a panel of 5–10 microsatellite markers recommended by the National Institute of Health [11]. MSI was defined as the presence of altered/additional bands in the polymerase chain reaction (PCR) amplified product of tumour DNA in comparison with the matched normal DNA samples obtained from the adjacent normal colon. Tumours were designated MSI-high if $\geq 40\%$ of the markers show altered band patterns, MSI-low if there is $< 40\%$ instability, and MSS if there was no instability [11].

MMR gene mutation analysis

MMR (*MSH2*, *MLH1*, *MSH6*) mutational analysis was performed on DNA from peripheral blood lymphocytes or lymphoblastoid cell lines. Large genomic deletions/duplications in *MSH2* and *MLH1* genes were identified by MLPA [12] and, if absent, underlying germline mutations were further screened by genomic DNA sequencing [9]. Briefly, the entire coding regions of the *MSH2* (16 exons) and *MLH1* (19 exons) genes were amplified by PCR and screened for mutations using ABI 377 automated sequencer. Similar functional analysis was also performed for *MSH6* gene using exon-by-exon sequencing strategy. Functional mutations were confirmed by assessing published literature, as well as from the human genome mutation database [13] and the International Collaborative Group-HNPCC database (InSIGHT). We used the computational programs Polymorphism Phenotyping to predict the pathogenicity of novel MMR alterations <http://genetics.bwh.harvard.edu/pph/> [14] and Sorting Intolerant From Tolerant <http://blocks.fhcrc.org/sift/SIFT.html> [15]. The description of the mutations found in this study is given in Table 1.

Statistical Methods

Kaplan-Meier survival analysis

For the cohort of proband's family members, Kaplan-Meier survival analysis [10] was used to estimate the age-specific cumulative CRC incidence, with the corresponding 95% confidence interval (CI), where observation time was taken from birth to the earlier of age at diagnosis of CRC or current age. The analysis was stratified by: (a) MMR mutation status (*MSH2* or *MLH1*); (b) gender among MMR mutation carriers and non-carriers.

Segregation-based analysis

To account for missing genotype data and the non-random ascertainment of the families, we fitted likelihood-based segregation analyses [16,17]. This allowed the estimation of the cumulative risk (penetrance) associated with either any Lynch Syndrome mutation or specific *MLH1* or *MSH2* mutations, in males and females. Based on preliminary data analysis, we found that a log-logistic regression model for the hazard function (see Equation (1) in Appendix 1) fitted the data well, and we used this to estimate the cumulative risk of developing CRC by each

Table 1: Description of MMR mutation-carrying participants – mutation information and counts of kin (family size), kin with blood, colorectal cancer (CRC) amongst kin, and mutation carriers amongst kin by relative degree

Gene	Germline Mutation	Mutation Consequence	No. of kin reported		No. of kin with CRC reported		No. of kin with blood		No. of mutation carriers	
			1 st	2 nd	1 st	2 nd	1 st	2 nd	1 st	2 nd
MLH1	c.1852_1853 AA>GC	p.Lys618Ala	4	16	2	0	2	5	0	0
	c.1689_1690insA	p.Ile563IlefsX4	3	12	1	2	2	0	0	0
	c.1732_1896del	Exon 16 del	6	11	2	0	2	4	1	3
	c.116+5 G>C	Splice-site defect	7	13	1	1	2	0	1	0
	c.793C>T	p.Arg265Cys	10	23	2	5	1	6	1	1
	c.793C>T	p.Arg265Cys	7	25	1	5	2	5	1	0
	c.346delA	p.Thr116GlnfsX20	2	5	1	2	0	0	0	0
	c.731G>A	p.Gly244Asp	6	5	2	0	5	0	5	0
	c.790+2T>C	Splice-site defect	9	19	3	1	6	7	3	3
	c.298C>T	p.Arg100X	7	6	2	0	1	0	0	0
	c.1975C>T	p.Arg659X	3	5	2	3	3	0	0	0
	c.2223_2231del	In-frame deletion	9	11	2	1	3	0	2	0
	c.793C>T	p.Arg265Cys	4	11	1	1	2	0	1	0
	c.350C>T	Thr117Met	9	10	1	0	4	0	3	0
	Total:			86	172	23	21	35	27	18
MSH2	c.1216C>T	p.Arg406X	5	5	2	0	1	1	0	0
	c.1165C>T	p.Arg389X	3	7	1	2	1	1	1	1
	c.1277_1386del	Exon 8 deletion	5	18	4	3	1	2	1	0
	c.2075G>T	p.Gly692Val	5	10	1	3	0	0	0	0
	c.136_164del	p.His46GlyfsX35	3	5	1	2	0	0	2	0
	c.2135_2136insT	p.Val712ValfsX4	5	11	2	1	1	0	0	0
	c.942+3A>T	Splice-site defect	13	33	3	1	1	3	1	0
	c.1511-2A>G	Splice-site defect	9	3	1	0	2	1	2	1
	c.363T>G	p.Tyr121X	10	23	3	1	7	1	1	0
	c.1_1386del	Exons 1–8 deletion	9	21	3	0	4	6	3	3
	c.1705_1706del	p.Glu569IlefsX2	3	7	1	2	1	4	1	3
	c.965G>T	p.Gly322Asp	6	14	1	0	1	0	0	0
	c.942+3A>T	Splice-site defect	5	9	2	1	2	1	2	1
	c.645+1G>A	Splice-site defect	5	7	1	2	1	0	1	0
	c.1705_1706del	p.Glu569IlefsX2	5	12	2	0	2	0	0	0
	c.1165C>T	p.Arg389X	7	22	3	1	3	6	2	3
c.1165C>T	p.Arg389X	3	5	1	1	1	2	1	2	
Total:			101	212	32	20	31	28	18	14
MSH6	c.3335_3336insA TGA	p.Asp1112GlnfsX2	12	37	1	0	5	1	2	1
Total:			12	37	1	0	5	1	2	1
Grand total:			199	421	56	41	71	56	38	22

(1st = first degree relatives, 2nd = second degree relatives)

decade of age, for each type of mutation. The log-logistic model had the added advantage that it allowed the hazard rate to follow a non-monotonic function, in which the risk initially increased with age and then decreased, as observed in Jenkins et al. [18].

The analyses were adjusted for ascertainment by conditioning the likelihood of the family's observed genotypes on the observed phenotypes in the family members, given their ages at examination (see Appendix 2). Thus, we obtain the maximum likelihood estimates of the needed

Table 2: Estimates of cumulative risks (%) to ages 30, 50, and 70 (and corresponding 95% confidence intervals) for colorectal cancer stratified by MMR mutation status and gender using Kaplan-Meier analysis

Mutation	Gender	Age 30	Age 50	Age 70
Carrier of any MMR mutation Carrier	Male (n = 27)	-	21 (0, 39)	67 (3, 89)
	Female (n = 33)	4 (0, 10)	35 (6, 55)	51 (16, 71)
	Combined (n = 60)	2 (0, 6)	29 (10, 44)	55 (28, 72)
<i>MLH1</i> carrier	Male (n = 13)	-	40 (0, 70)	70 (0, 94)
	Female (n = 12)	-	-	38 (0, 68)
	Combined (n = 25)	-	16 (0, 34)	50 (4, 73)
<i>MSH2</i> carrier	Male (n = 12)	-	-	-
	Female (n = 23)	5 (0, 15)	61 (9, 83)	61 (9, 83)
	Combined (n = 35)	3 (0, 10)	36 (6, 56)	56 (15, 77)
Non-carrier of all mutations	Male (n = 22)	-	-	-
	Female (n = 40)	-	7 (0, 19)	7 (0, 19)
	Combined (n = 62)	-	4 (0, 11)	4 (0, 11)
Any mutation unknown	Male (n = 195)	2 (0, 5)	23 (14, 31)	46 (34, 57)
	Female (n = 157)	4 (0, 7)	13 (6, 20)	41 (27, 52)
	Combined (n = 352)	3 (1, 5)	19 (13, 24)	44 (35, 52)

(n represents the number of kin with blood available and age information available)

parameters by maximizing this ascertainment-corrected retrospective likelihood [16,17] into the software Mendel [19] and used the parameters to estimate the cumulative risk of CRC at a given age, using the penetrance function in Equation (2) of the Appendix 1. For this analysis, the ascertainment correction was based on the first degree relatives of probands (i.e., the proband, his/her sibs and their parents), where probands carried a MMR mutation. For all analyses, the 95% CI for the cumulative risk by a given age was estimated using simulation of 1,000 sets of parameters, assuming a multivariate normal distribution of the parameters being estimated [18].

Results

Distribution of missing ages and ages at onset in the families

We had information on age at CRC onset for all 32 probands. This information was missing for 6 out of 97 affected relatives but 3 of them had their age at examina-

tion available. Among the 493 unaffected individuals (255 males and 238 females), age at examination was missing for 113 (60 males and 53 females). Among affected individuals (probands and relatives), the mean age at onset was 45.1 (n = 25, standard deviation (s.d.) = 13.3) in *MSH2* carriers, 53.6 (n = 21, s.d. = 12.3) among *MLH1* carriers and 51.4 (n = 76, s.d. = 14.6) for individuals with unknown mutation status. Only one individual was negative for *MSH2* among the affected relatives.

Kaplan-Meier analysis

Descriptive analyses with the Kaplan-Meier (KM) estimator were carried out on the sample of kin with blood and age information available. The kin included 60 individuals who were positive for any MMR mutation (16 affected and 44 unaffected individuals) and 62 who were negative (1 affected and 61 unaffected individuals). In addition, we also estimated separately the cumulative risk in 352 individuals with unknown mutation status (77 affected

Table 3: Estimates of cumulative risks (%) at ages 30, 50 and 70 (and corresponding 95% confidence intervals) for colorectal cancer among mutation carriers and non-carriers (any MMR, MLH1, MSH2) based on the segregation analysis.

Mutation	Gender	Age 30	Age 50	Age 70
Carrier of any MMR mutation	Male	1 (0, 2)	23 (10, 33)	60 (35, 73)
	Female	1 (0, 1)	15 (7, 23)	47(27, 60)
	Combined	1 (0, 1)	18 (11, 25)	53 (37, 64)
MLH1 Carrier	Male	1 (0, 2)	25 (7, 51)	67 (27, 89)
	Female	0 (0, 1)	8 (2, 19)	35 (10, 59)
	Combined	0 (0, 12)	13 (4, 30)	44 (19, 70)
MSH2 Carrier	Male	1 (0, 2)	21 (1, 36)	55 (2, 75)
	Female	1 (0, 2)	20 (1, 32)	53 (2, 70)
	Combined	1 (0, 2)	20 (1, 29)	54 (3, 69)
Non-carrier of all mutations	Male	0 (0, 1)	2 (0, 7)	10 (1, 27)
	Female	0 (0, 1)	1 (0, 6)	6 (1, 24)
	Combined	0 (0, 1)	2 (0, 6)	9 (2, 24)

Figure 1
Age-specific cumulative risks of colorectal cancer (CRC) among carriers and non-carriers of MMR (any MMR, MLH1, MSH2) mutation specified by gender, based on the segregation analysis.

Figure 2
Gender-specific and/or mutation-specific hazard rate estimates of developing colorectal cancer (CRC) in *MLH1* gene (left) and *MSH2* gene (right) mutations, based on the segregation analysis.

and 275 unaffected individuals). By age 70, about two thirds of male carriers and one half of female carriers of any MMR mutation had experienced CRC (Table 2). Importantly, the risk among those with unknown mutation status was considerably closer to that of carriers than non-carriers between the two estimates (Table 2). Thus, omitting the unknown group or combining their estimates with another genotype category could lead to serious biases.

Modified segregation-based analysis

The modified segregation-based approach was able to use the full cohort of probands and their kin, with or without blood available. After removing individuals without information on age at diagnosis or examination, the final sample included 506 individuals out of the 622 recruited. Therefore, the estimation of the cumulative risk with the segregation approach took into account 32 probands, 60 kin who tested positive for any MMR mutation, 62 kin who tested negative, and 352 kin with unknown mutation status. These 32 probands and 352 kin could not be included in the KM estimation of carriers and non-carriers.

Table 4: Estimated hazard ratios of colorectal cancer risk at ages 30, 50 and 70 (and corresponding 95% confidence interval) in gene mutation carriers (any MMR, *MLH1*, *MSH2*) compared with that in the general population, based on the segregation analysis.

Mutation	Age 30	Age 50	Age 70
Any MMR	10.5 (2.5, 46.9)	8.8 (2.3, 40.1)	5.5 (1.8, 25.0)
<i>MLH1</i>	5.5 (0.8, 53.6)	6.5 (1.4, 41.8)	5.1 (1.3, 26.9)
<i>MSH2</i>	13.1 (0.3, 78.8)	9.3 (0.3, 44.4)	5.4 (0.3, 24.5)

As assessed by segregation analysis, 60% of male and 47% of female carriers of any MMR gene mutation had developed CRC by age 70. These risks were 81% and 72% by age 90, respectively. The cumulative risk appeared higher for men than for women at all ages (Table 3 and Figure 1), although a formal test of difference was not significant (Wald Chi-square test; p-value = 0.25). This arose because male carriers of *MLH1* gene mutations had a significantly higher risk of cancer than female carriers at all ages (67% vs. 35% by age 70, p-value = 0.02) (Table 3 and Figure 1), while male and female carriers of *MSH2* mutations had similar cumulative risks of disease (\approx 54%, p-value = 0.89) (Table 3). Among men, the estimated risk was higher among carriers of *MLH1* mutations than *MSH2* mutations (e.g. 67% vs. 55% by age 70), while the opposite was observed among women (e.g. 35% vs. 53% by age 70).

For *MLH1* mutations, the age-specific risk of cancer among male carriers (the hazard rate) increased to about age 70 but remained relatively constant thereafter (Figure 2), while female carriers continued to experience an increase in risk after age 70. However, for *MSH2* mutation carriers of both sexes, risk increased only until about age 70 and stabilized afterwards. As observed in previous studies, male carriers were at higher risk than female carriers for *MLH1* but not for *MSH2* mutations.

Estimates of hazard ratios (HRs) for CRC at different ages are presented in Table 4. Estimates for males and females were combined in the table due to their large confidence intervals. The effect of *MLH1* on CRC was almost constant with age (the HR varied between 5.5 at age 30 and 5.1 at age 70). *MSH2* had a stronger effect on CRC and exhibited a decreasing pattern with age (the HR fell from 13.1 at age 30 to 5.4 at age 70). The global effect of any MMR muta-

tion on CRC was significant and showed a decreasing trend with age (the HR fell from 10.5 at age 30 to 5.5 at age 70).

Discussion

To our knowledge, only three other studies [6,18,20] have provided population-based estimate of CRC risk in Lynch Syndrome families but none of them was reported in North America. While our results confirmed a relatively high penetrance associated with MMR gene mutations, our risk estimates seem lower than many clinic-based estimates [3,5,21-26].

Previous studies have estimated the risk of developing CRC among MMR gene mutation carriers in Lynch Syndrome families to vary between 30 to 100% [3-6,21-26], where the lowest rates are generally reported in women from population-based studies. The excess of risk in males compared to females found in several studies [3,4,6] was also confirmed in our analyses. This could suggest that females are protected from CRC; perhaps due to environmental/reproductive factors unique to women or to a sex-linked modifier gene. We also found a gender specific mutation effect with a higher risk in male carriers of *MLH1* mutations (67%) vs. carriers of *MSH2* mutations (55%), while the opposite was observed among women carriers (35% for *MLH1* vs. 53% for *MSH2*), by age 70. Our study also showed that age-specific risk (hazard ratio) associated with *MLH1* was almost constant with age, while the risk associated with *MSH2* decreased with age. The same trends were also suggested in a recent study [18]. If differences in risks observed for *MLH1* and *MSH2* are confirmed, then the distribution of the types of MMR mutations could have a profound impact on the cancer risk.

Our cumulative risk estimates are close to the lower estimates previously published. They are slightly higher than another recent population-based study [18], but this latter selected exclusively early-onset probands. Several factors might explain the discrepancies with the previous penetrance estimates. First, methods of kindred ascertainment varied between studies, and the distribution of factors which likely affect risk may also vary across studies conducted in different countries. Because MMR mutations are rare in the general population [6,8], most penetrance estimates are derived from high-risk Lynch Syndrome families, who usually satisfy the original or revised Amsterdam criteria [27]. Because these designs are enriched with mutation carriers, they could be more efficient for estimating penetrance than population-based designs [17] but also more prone to biases [28,29]. Extrapolation to the general population (i.e. all CRC cases) is not possible unless appropriate ascertainment correction is applied to

account for the nonrandom sampling. This was performed in this study by the use of the modified segregation-based analysis. Our recent simulation studies [17] confirmed the validity of our ascertainment-corrected approach. Second, it is likely that there exist other genetic and non-genetic contributors to HNPCC, other than a MMR mutation, that could also aggregate within families. Some of our additional analyses suggest the role of a second major gene within these families, but further work is still needed to distinguish its effect from a common environmental factor. Third, data quality was improved in the present investigation compared to previous studies in two ways. Previous studies, unlike the present study, were conducted before techniques were available to test for MMR gene mutations so could not determine accurately the status of the MMR gene [22,23]. In addition, missing genotype is a common problem in family studies and the classical analysis approaches for time-to-onset data such as Kaplan-Meier estimation or Cox regression model, in their original formulations, cannot solve this problem. In this study, the use of a modified segregation-based approach allows inferences on the missing genotypes by using the Mendelian transmission probabilities and genealogical relationships. As a consequence, the segregation-based analysis was able to use the information on 32 probands and 352 kin who were not included in the KM analyses, resulting in more precise and potentially less biased penetrance estimates.

Our estimated cumulative risks among non-carriers are much higher than observed in the general population, for example we estimated a combined (*MLH1/MSH2*) cumulative risk of 9% by age 70 while it is about 2% in the US population. The possible discrepancy between our cumulative risk estimates in non-carriers vs. those published in the US general population could be due to the fact that our sample is enriched with affected individuals who could have a very different genetic and non-genetic risk profile than the general population. Therefore, even if our design and ascertainment correction approach tend to yield estimates that are closer to the general population, the difference seen between the two estimates can reflect a difference in the distribution of risk factors in our sample compared to the general population [28].

In summary, many sources of bias have been reduced in this study, in part through the choice of study subjects and use of the modified segregation-based analyses. However, several limitations may still exist. First, the sample size is relatively small for a risk estimation study and the confidence intervals are large, especially for estimating mutation-specific and gender-specific penetrances. This problem was only partly overcome by using the modified segregation-based approach. Efficiency was also improved

by selecting preferentially probands carrying a mutation and thus they are more informative than random probands [17]. Second, inaccuracy of cancer history might introduce error. While we attempted to confirm with pathology reports all reported family members with a CRC diagnosis, this was not possible for all cases. However, recent research conducted in Ontario found that proband's reports of relatives' cancer diagnoses are fairly accurate; 93% of proband-reported CRC among first-degree relatives were verified by either hospital records, cancer registry or death certificates, though reporting was less accurate for second-degree relatives with only 72% of reported CRCs verified [30]. Third, some studies classified tumours as MSI-high if > 30% of the markers show altered band patterns. Because we identified MSI-high tumours as those with altered band patterns in > 40% of markers, we may have missed some carriers. However, because this classification was not associated with the proband's family history, it is unlikely that this biased our penetrance estimates. Finally, we assumed that the probands selected are representative of the entire population of CRC families in Ontario. Because of our relatively small sample size, this hypothesis is difficult to assess and deviation from this hypothesis could lead to a selection bias. Although unlikely, it is possible that the estimates of penetrance could be biased upwards if families carrying the gene participated differentially according to the prevalence of cancer in the family.

In conclusion, this study provides a unique population-based study of CRC risks among *MSH2/MLH1* mutation carriers in a Canadian population and can help to better define and understand the patterns of risks among members of Lynch Syndrome families.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

YC, LB performed the statistical developments and analyses and drafted the manuscript. MC conceived of the study, participated in its design and coordination and contributed to the manuscript drafting. BP carried out the molecular genetic studies. MN contributed to the statistical analysis. KB participated in the study coordination and data collection. SG, JML, GME, MA participated in the study design and manuscript drafting. All authors read and approved the final manuscript.

Appendix 1: Log-logistic regression model

We consider a log-logistic regression model for investigating the dependence of age at onset of colorectal cancer on mutation carrier status and gender.

Suppose T is the age of onset associated with genotype G and other covariates x . We assume T follows the log-logistic distribution, which has hazard function $h(t|G)$ of the form by

$$h(t | G, x) = \frac{\lambda \rho \{ \lambda(t-20) \}^{\rho-1} e^{\beta_s x_s + \beta_1 x_1}}{1 + \{ \lambda(t-20) \}^{\rho} e^{\beta_s x_s + \beta_1 x_1}}, \quad (1)$$

where x_1 indicates the mutation carrier status for the MMR gene (either *MLH1*, *MSH2*, or any) and x_s distinguishes between males ($x_s = 1$) and females ($x_s = 2$). Here, the parameter $\lambda > 0$ is a scale parameter and $\rho > 0$ is a shape parameter, which allows a non-monotonic hazard function: when $\rho > 1$, the hazard function is unimodal and when $\rho \leq 1$, it decreases monotonically.

Then, the corresponding penetrance function is defined by the cumulative risk of a disease up to age t given the genotype G and other covariates x , as

$$F(t | G, x) = \frac{\{ \lambda(t-20) \}^{\rho} e^{\beta_s x_s + \beta_1 x_1}}{1 + \{ \lambda(t-20) \}^{\rho} e^{\beta_s x_s + \beta_1 x_1}}. \quad (2)$$

Appendix 2: Ascertainment-corrected retrospective likelihood

Our risk estimates (relative or cumulative) are based on an ascertainment-corrected retrospective likelihood approach. Correction was based on the conditional probability of the observed genotypes (G) given the observed phenotypes (D) in the family and corrected for the ascertainment event (A).

The ascertainment-corrected retrospective likelihood, L , arising from a sample of n independent families for each family size n_f has the form

$$L = \prod_{f=1}^n L_f = \prod_{f=1}^n \frac{N_f}{A_f},$$

where L_f is the conditional likelihood of family f obtained by dividing the likelihood contribution of family f , N_f , by the probability of its being ascertained, A_f .

For family f with n_f family members, we define $D = (D_1, \dots, D_{n_f})$ and $G = (G_1, \dots, G_{n_f})$ as the vector forms that represent their phenotypes and genotypes, respectively, where the phenotype is defined by the age at onset and affection status, i.e. $D_i = (T_i, \delta_i)$. The likelihood contribution for family f can be expressed as a function of the genotypes of family members given their phenotypes and ascertainment event, i.e.,

$$L_f = P(G | D, A) \propto \frac{P(D|G)P(G)}{P(D,A)}$$

Here, $P(D|G)$ in the numerator can be obtained using survival analysis approach,

$$P(D | G) = \prod_{i=1}^{n_f} h(t_i | G_i, x_i)^{\delta_i} S(t_i | G_i, x_i),$$

where $h(\cdot)$ and $S(\cdot)$ represent hazard and survivor functions, respectively, defined in Appendix 1, $P(G)$ is the genotype probability of family members based on Hardy-Weinberg equilibrium and Mendelian transmission probabilities, and the denominator, $P(D, A)$, represents the probability of observing the phenotypes of individuals who qualify for ascertainment. For more details, refer Choi et al. [17].

Acknowledgements

We would like to thank the Ontario Familial Colorectal Cancer Registry staff, in particular Darshana Daftary and Kevin Boyd (Cancer Care Ontario), and Teresa Selander (Mount Sinai Hospital), the biospecimen manager, for their dedication to this study. The content of this manuscript does not necessarily reflect the views or policies of the NCI, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the Consortium of Familial Registries.

This research was supported by the Canadian Institutes of Health Research-Interdisciplinary Health Research Team, grant no. CRT 43821 (M. Cotterchio, M. Neerav, B. Bapat, K. Boyd, S. Gallinger, J. McLaughlin); The National Cancer Institute, the National Institutes of Health, RFA # CA-95-011, grant no. U01-CA74783; The Institutes of Genetics and Population and Public Health of the Canadian Institutes of Health Research, Grant no. 110053 (L. Briollais, Y.-H. Choi, J. McLaughlin); The Canadian Breast Cancer Foundation-Ontario Chapter (Y.-H. Choi).

References

- Lynch HT, Watson P, Shaw TG, Lynch JF, Harty AE, Franklin BA, Kapler CR, Tinley ST, Liu B, Lerman C: **Clinical impact of molecular genetic diagnosis, genetic counseling, and management of hereditary cancer. Part II: Hereditary nonpolyposis colorectal carcinoma as a model.** *Cancer* 1999, **86**(11 Suppl):2457-2463.
- Bocker T, Ruschoff J, Fishel R: **Molecular diagnostics of cancer predisposition: HNPCC and mismatch repair defects.** *Biochimica Biophysica Acta* 1999, **1423**:1-10.
- Aarnio M, Sankila R, Pukkala E, Salovaara R, Aaltonen LA, de la Chapelle A, Peltomäki P, Mecklin J-P, Järvinen H: **Cancer risk in mutation carriers of DNA-mismatch-repair genes.** *Int J Cancer* 1999, **81**:214-218.
- Lin KM, Shashidharan M, Thorson AG, Ternent CA, Blatchford GJ, Christensen MA, Watson P, Lemon SJ, Franklin B, Karr B, Lynch J, Lynch HT: **Cumulative incidence of colorectal and extracolonic cancers in *MLH1* and *MSH2* mutation carriers of hereditary nonpolyposis colorectal cancer.** *J Gastrointest Surg* 1998, **2**:67-71.
- 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: **Cancer risk in families with hereditary nonpolyposis colorectal cancer diagnosed by mutation analysis.** *Gastroenterology* 1996, **110**:1020-1027.
- Dunlop MG, Farrington SM, Carothers AD, Wyllie AH, Sharp L, Burn J, Liu B, Kinzler KW, Vogelstein B: **Cancer risk associated with germline DNA mismatch repair gene mutations.** *Hum Mol Genet* 1997, **6**:105-110.
- Cotterchio M, McKeown-Eyssen G, Sutherland H, Buchan G, Aronson M, Easson AM, Macey J, Holowaty E, Gallinger S: **Ontario familial colon cancer registry: methods and first year response rates.** *Chron Dis Canada* 2000, **21**:81-86.
- Dunlop MG, Farrington SM, Nicholl I, Aaltonen L, Petersen G, Porteous M, Carothers A: **Population carrier frequency of *hMSH2* and *hMLH1* mutations.** *Br J Cancer* 2000, **83**:1643-5.
- Woods MO, Hyde AJ, Curtis FK, Stuckless S, Green JS, Pollett AF, Robb JD, Green RC, Croitoru ME, Careen A, Chaulk JAW, Jegathesan J, McLaughlin JR, Gallinger SS, Younghusband HB, Bapat BV, Parfrey PS: **High frequency of hereditary colorectal cancer in Newfoundland likely involves novel susceptibility genes.** *Clin Cancer Res* 2005, **11**:6853-61.
- Kaplan EL, Meier P: **Nonparametric estimation from incomplete observations.** *J Amer Statist Assoc* 1958, **53**:457-481.
- Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani N, Srivastava S: **A national Cancer Institute workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer.** *Cancer Res* 1998, **58**:5248-5257.
- Bapat BV, Madlensky L, Temple LKF, Hiruki T, Redston M, Baron DL, Ling X, Marcus VA, Soravia C, Mitri A, Shen W, Gryfe R, Berk T, Chodirker BN, Cohen Z, Gallinger S: **Family history characteristics, tumour microsatellite instability and germline *MSH2*, and *MLH1* mutations in hereditary colorectal cancer.** *Hum Genet* 1999, **104**:167-176.
- Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G: **Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification.** *Nucleic Acids Res* 2002, **30**:e57.
- Ramensky V, Bork P, Sunyaev S: **Human non-synonymous SNPs: server and survey.** *Nucleic Acids Res* 2002, **30**:3894-3900.
- Ng PC, Henikoff S: **Predicting deleterious amino acid substitutions.** *Genome Res* 2001, **11**:863-74.
- Kraft P, Thomas DC: **Bias and efficiency in family-based gene-characterization studies: conditional, prospective, retrospective, and joint likelihoods.** *Am J Hum Genet* 2000, **66**:1119-1131.
- Choi YH, Kopciuk K, Briollais L: **Estimating disease risk associated with mutated genes in family-based designs.** *Hum Hered* 2008, **66**:238-251.
- Jenkins MA, Baglietto L, Dowty JG, Van Vliet CM, Smith L, Mead LJ, Macrae FA, St John DJ, Jass JR, Giles GG, Hopper JL, Southey MC: **Cancer risks for mismatch repair gene mutation carriers: a population-based early onset case-family study.** *Clin Gastroenterol Hepatol* 2006, **4**:489-98.
- Lange K, Sinsheimer JS, Sobel E: **Association testing with Mendel.** *Genet Epidemiol* 2005, **29**:36-50.
- Quehenberger F, Vasen HFA, van Houwelingen HC: **Risk of colorectal and endometrial cancer for carriers of mutations of the *hMLH1* and *hMSH2* gene: correction for ascertainment.** *J Med Genet* 2005, **42**:491-496.
- Green J, O'Driscoll M, Barnes A, Maher ER, Bridge P, Shields K, Parfrey PS: **Impact of gender and parent of origin on the phenotypic expression of hereditary nonpolyposis colorectal cancer in a large Newfoundland kindred with a common *MSH2* mutation.** *Dir Colon Rectum* 2002, **45**:1223-1232.
- Aarnio M, Mecklin JP, Aaltonen LA, Nystrom-Lahti M, Jarvinen HJ: **Life-time risk of different cancers in HNPCC syndrome.** *Int J Cancer* 1995, **64**:430-433.
- Froggart NJ, Green J, Brassett C, Evans E, Bishop D, Kolodner R, Maher E: **A common *MSH2* mutation in English and North American HNPCC families: origin, phenotypic expression, and sex specific differences in colorectal cancer risk.** *J Med Genet* 1999, **36**:97-102.
- Vasen HFA, Stormorken A, Menko FH, Nagengast FM, Kleibeuker JH, Griffioen G, Taal BG, Moller P, Wijnen JT: ***MSH2* mutation carriers are at higher risk of cancer than *MLH1* mutation carriers: a study of hereditary nonpolyposis colorectal cancer families.** *J Clin Oncol* 2001, **19**:4074-4080.

25. Parc Y, Boisson C, Thomas G, Olschwang S: **Cancer risk in 348 French MSH2 or MLH1 gene carriers.** *J Med Genet* 2003, **40**:208-213.
26. Hampel H, Stephens JA, Pukkala E, Sankila R, Aaltonen LA, Mecklin JP, de la Chapelle A: **Cancer risk in hereditary nonpolyposis colorectal cancer syndrome: later age of onset.** *Gastroenterology* 2005, **129**:415-21.
27. Vasen HFA, Watson P, Mecklin JP, Lynch HT: **New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch Syndrome) proposed by the International Collaborative Group on HNPCC.** *Gastroenterology* 1999, **116**:1453-1456.
28. Begg CB: **On the use of familial aggregation in population-based case probands for calculating penetrance.** *J Nat Cancer Inst* 2002, **94**:1221-1226.
29. Carayol J, Bonāiti-Pellié C: **Estimating penetrance from family data using a retrospective likelihood when ascertainment depends on genotype and age of onset.** *Gen Epidemiol* 2004, **27**:109-117.
30. Theis B, Boyd N, Lockwood G, Tritchler D: **Accuracy of family cancer history in breast cancer patients.** *Eur J Cancer Prev* 1994, **3**:321-327.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

