Author/s:
Win, AK;Dowty, JG;Cleary, SP;Kim, H;Buchanan, DD;Young, JP;Clendenning, M;Rosty, C;Macinnis, RJ;Giles, GG;Boussioutas, A;Macrae, FA;Parry, S;Goldblatt, J;Baron, JA;Burnett, T;Le Marchand, L;Newcomb, PA;Haile, RW;Hopper, JL;Cotterchio, M;Gallinger, S;Lindor, NM;Tucker, KM;Winship, IM;Jenkins, MA

Title:
Risk of colorectal cancer for carriers of mutations in MUTYH, with and without a family history of cancer

Date:
2014-01-01

Citation:
Win, A. K., Dowty, J. G., Cleary, S. P., Kim, H., Buchanan, D. D., Young, J. P., Clendenning, M., Rosty, C., Macinnis, R. J., Giles, G. G., Boussioutas, A., Macrae, F. A., Parry, S., Goldblatt, J., Baron, J. A., Burnett, T., Le Marchand, L., Newcomb, P. A., Haile, R. W. ,... Jenkins, M. A. (2014). Risk of colorectal cancer for carriers of mutations in MUTYH, with and without a family history of cancer. Gastroenterology, 146 (5), pp.1208-12e5. https://doi.org/10.1053/j.gastro.2014.01.022.

Persistent Link:
https://hdl.handle.net/11343/43899

Risk of Colorectal Cancer for Carriers of Mutations in *MUTYH*, with and without a Family History of Cancer

Aung Ko Win[1], James G. Dowty[1], Sean P. Cleary[2,3], Hyeja Kim[2,3], Daniel D. Buchanan[4], Joanne P. Young[4], Mark Clendenning[4], Christophe Rosty[4,5,6], Robert J. MacInnis[1,7], Graham G. Giles[1,7], Alex Boussioutas[8,9,10], Finlay A. Macrae[11], Susan Parry[12,13], Jack Goldblatt[14], John A. Baron[15], Terrilea Burnett[16], Loïc Le Marchand[16], Polly A. Newcomb[17], Robert W. Haile[18], John L. Hopper[1], Michelle Cotterchio[3], Steven Gallinger[2,3], Noralane M. Lindor[19], Katherine M. Tucker[20], Ingrid M. Winship[8,21], Mark A. Jenkins[1]

Running Title

Cancer risks for *MUTYH* mutation carriers

Affiliations

[1] Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, The University of Melbourne, Parkville, Victoria, Australia.
[2] Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada.
[3] Cancer Care Ontario, Toronto, Ontario, Canada.
[4] Cancer and Population Studies Group, Queensland Institute of Medical Research, Bancroft Centre, Herston, Queensland, Australia.
[5] Department of Molecular and Cellular Pathology, University of Queensland, Herston, Queensland, Australia.
[6] Envoi Specialist Pathologists, Herston, Queensland, Australia.
[7] Cancer Epidemiology Centre, Cancer Council Victoria, Carlton, Victoria, Australia.
[8] Department of Medicine, The University of Melbourne, Parkville, Victoria, Australia.
[9] Cancer Genomics and Predictive Medicine, Peter MacCallum Cancer Centre, East Melbourne, Victoria, Australia.
[10] Sir Peter MacCallum Department of Oncology, The University of Melbourne, Parkville, Victoria, Australia.
[11] Colorectal Medicine and Genetics, The Royal Melbourne Hospital, Parkville, Victoria, Australia.
[12] New Zealand Familial Gastrointestinal Cancer Registry, Auckland City Hospital, Auckland, New Zealand.
[13] Department of Gastroenterology, Middlemore Hospital, Auckland, New Zealand New Zealand.
[14] Genetic Services & Familial Cancer Program of Western Australia and School of Paediatrics and Child Health, University of Western Australia, Perth, Australia
[15] Department of Medicine, University of North Carolina, Chapel Hill, North Carolina, USA.
[16] University of Hawaii Cancer Center, Honolulu, Hawaii, USA.
[17] Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.
[18] Department of Medicine, Division of Oncology, Stanford University, California, USA.
[19] Department of Health Science Research, Mayo Clinic Arizona, Scottsdale, Arizona, USA.
[20] Hereditary Cancer Clinic, Prince of Wales Hospital, Randwick, New South Wales, Australia.

[21] Genetic Medicine and Family Cancer Clinic, Royal Melbourne Hospital, Parkville, Australia

Corresponding author
Mark Jenkins, PhD
Centre for Molecular, Environmental, Genetic and Analytic Epidemiology
Melbourne School of Population and Global Health
Level 3, 207 Bouverie Street
The University of Melbourne
VIC 3010 Australia
Tel: +61 3 8344 0902; Fax: +61 3 9349 5815
Email: m.jenkins@unimelb.edu.au

FUNDING

This work was supported by the National Cancer Institute, National Institutes of Health under RFA #CA-95-011 and through cooperative agreements with members of the Colon Cancer Family Registry and Principal Investigators. Collaborating centers include Australasian Colorectal Cancer Family Registry (U01 CA097735), Familial Colorectal Neoplasia Collaborative Group (U01 CA074799) [USC], Mayo Clinic Cooperative Family Registry for Colon Cancer Studies (U01 CA074800), Ontario Registry for Studies of Familial Colorectal Cancer (U01 CA074783), Seattle Colorectal Cancer Family Registry (U01 CA074794), and University of Hawaii Colorectal Cancer Family Registry (U01 CA074806). This work was also supported by a grant from the National Cancer Institute, USA (R01CA170122) and a Centre for Research Excellence grant from the National Health and Medical Research Council (NHMRC), Australia (APP1042021). AKW is supported by the Picchi Brothers Foundation Cancer Council Victoria Cancer Research Scholarship, Australia. MAJ is an NHMRC Senior Research Fellow. JLH is a NHMRC Senior Principal Research Fellow. CR is a Jass Pathology Fellow.

DISCLAIMER

The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centers in the CFRs, nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government or the CFR. Authors had full responsibility for the design of the study, the collection of the data, the analysis and interpretation of the data, the decision to submit the manuscript for publication, and the writing of the manuscript.

DISCLOSURE

The authors have no conflict of interest to declare with respect to this manuscript.

ACKNOWLEDGEMENTS

The authors thank all study participants of the Colon Cancer Family Registry and staff for their contributions to this project.

Authors’ Contributions

Aung Ko Win: study concept and design; acquisition of data; statistical analysis; interpretation of data; drafting and critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

James G. Dowty: study concept and design; statistical analysis; interpretation of data; drafting and critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Sean P. Cleary: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Hyeja Kim: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Daniel D. Buchanan: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Joanne P. Young: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Mark Clendenning: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Christophe Rosty: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Robert J. MacInnis: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Graham G. Giles: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Alex Boussioutas: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Finlay A. Macrae: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Susan Parry: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Jack Goldblatt: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

John A. Baron: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Terrilea Burnett: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Loïc Le Marchand: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Polly A. Newcomb: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Robert W. Haile: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

John L. Hopper: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Michelle Cotterchio: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Steven Gallinger: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Noralane M. Lindor: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Katherine M. Tucker: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Ingrid M. Winship: acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

Mark A. Jenkins: study concept and design; acquisition of data; interpretation of data; critical review of the manuscript for important intellectual content; approval of the final version of the manuscript

ABSTRACT

We studied 2332 individuals with monoallelic mutations in *MUTYH* among 9504 relatives of 264 colorectal cancer (CRC) cases with a *MUTYH* mutation. We estimated CRC risks, through 70 y of age, of 7.2% for male carriers of monoallelic mutations (95% confidence interval [CI], 4.6%–11.3%) and 5.6% for female carriers of monoallelic mutations (95% CI, 3.6%–8.8%), irrespective of family history. For monoallelic *MUTYH* mutation carriers with a first-degree relative with CRC, diagnosed by 50 y of age who does not have the *MUTYH* mutation, risks of CRC were 12.5% for men and (95% CI, 8.6%–17.7%) and 10% for women (95% CI, 6.7%–14.4%). Risks of CRC for carriers of monoallelic mutations in *MUTYH* with a first-degree relative with CRC are sufficiently high to warrant more intensive screening than for the general population.

KEYWORDS: colon cancer, genetics, base excision repair gene, DNA damage response

MUTYH is a base excision repair gene that detects and protects against oxidative DNA damage.[1] Individuals with germline mutations in both alleles (biallelic mutation carriers), whether they are homozygotes or compound heterozygotes, develop *MUTYH*-associated polyposis, an autosomal recessive disorder with substantially increased risk of CRC.[2] Individuals with germline mutations in one allele (monoallelic mutation carriers) have a small increased risk of CRC[3-5]. Due to the rarity of these mutations,[4, 6] previous studies have had limited ability to provide precise estimates of age- and sex-specific CRC risks for *MUTYH* mutation carriers. Further, the variability in CRC risk between carriers has not been quantified. Modelling of this variability can indicate a potential role for modifiers of risk.

RESULTS

We identified 9504 relatives (4613 females) from the families of the 264 (236 population-based and 28 clinic-based) probands with a monoallelic or biallelic *MUTYH* mutation from the Colon Cancer Family Registry; 138 (52%) from USA, 81 (31%) from Canada, and 45 (17%) from Australia and New Zealand. In the relatives, we observed 261 CRCs (114 females) whose ages at diagnosis had a median of 65 (range 26-98) years. *MUTYH* mutation status was known for 340 relatives (13 biallelic mutation carriers, 142 monoallelic mutation carriers, and 185 non-carriers). We estimated an additional 43 biallelic and 2190 monoallelic mutation carriers among non-genotyped relatives, giving a total estimated number of 56 biallelic and 2332 monoallelic mutation-carrying relatives in our sample.

Our methods allowed for CRC risk estimation in mutation families to be due to the *MUTYH* mutation as well as polygenic factors (combination of a large number of CRC-associated genetic susceptibility loci).[7] We estimated CRC risks, through 70y of age, for male and female to be: 75.4% (95%CI, 41.2%–96.6%) and 71.7% (95%CI, 44.5%–92.1%), respectively, for biallelic mutation carriers, and 7.2% (95%CI, 4.6%–11.3%) and 5.6% (95%CI, 3.6%–8.8%), respectively, for monoallelic mutation carriers (Figure 1). The estimated CRC risks, through 70y of age, for monoallelic mutation carriers with a first-degree relative with CRC were similar whether the relative was untested or a non-carrier or a monoallelic mutation carrier: approximately 12% (95%CI, 9%–18%) and 10% (95%CI, 7%–14%) respectively for males and females in comparison with males and females from the general population (2.9% and 2.1% respectively). However, if their affected first-degree relative was a biallelic mutation carrier then risks of CRC, through 70y of age, for monoallelic mutation carriers was estimated to be 10.4% (95%CI, 7.0%–15.0%) and 8.2% (95%CI, 5.4%–12.0%) respectively for males and females (Table 1). In addition, we estimated CRC risks for six other scenarios (Supplementary Figure 1). The highest risk of CRC for a monoallelic mutation carrier corresponded to having two affected first-degree relatives: one is a biallelic mutation carrier and one is a non-carrier (Supplementary Figure 1C).

We found no evidence for a difference in hazard ratios of CRC for biallelic mutation carriers between males and females (108 (95%CI, 25.9–454) vs 129 (95%CI, 43.7–380); p=0.85), nor for monoallelic mutation carriers between males and females (2.46 (95%CI, 1.54–3.93) vs 2.67 (95%CI, 1.67–4.26); p=0.81). Hazard ratio of CRC for Y179C monoallelic carriers was higher than for G396D monoallelic carriers (4.81 (95%CI, 3.00–7.71) vs 2.42 (95%CI, 1.48–3.98); p=0.05), but there was no difference between biallelic carriers of Y179C and G396D (p=0.84) (Supplementary Table 1).

The standard deviation of the polygenic component was estimated to be 1.11 (0.74–1.49, $p<0.001$); see the Materials and Methods for a general formula relating this standard deviation to the hazard ratio. At ages less than 50y this formula reduces to Pharoah's formula for early-onset disease[7] and says that monoallelic *MUTYH* mutation carriers with an affected first-degree relative have CRC incidences approximately 4.58 (for males) or 4.97 (for females) times the population incidences. However, Supplementary Figure 2 gives precise hazard ratios for all ages and shows that by age 70y, Pharoah's formula over-estimates relative risks by roughly 30%.

DISCUSSION

Our finding of almost complete penetrance for biallelic *MUTYH* mutation carriers is consistent with previous studies.[8-10] There is some evidence that biallelic mutation carriers move rapidly along a mutator phenotype progression to cancer.[11] These findings support the recommendation that biallelic mutation carriers should consider prophylactic total colectomy with ileorectal anastomosis depending on the individual, age of presentation and number and size of polyps present.[12]

We estimated monoallelic mutation carriers had on average, an approximately 2.5-fold increased risk of CRC compared with the general population, consistent with one previous study.[13] This level of increased risk for monoallelic mutation carriers is similar to that for people with a first-degree relative with CRC, who are recommended 5-yearly colonoscopy starting 10y younger than the youngest case in the family and before age 50y.[13] However, monoallelic mutation carriers who have an affected first-degree relative were at approximately 5-fold increased risk. For these carriers, colonoscopy beginning at age 40y, with follow-up at intervals dependent on the presence or absence of polyps but no less often than every 5 years, may be reasonable.

We observed strong evidence that CRC risks for carriers are highly heterogeneous. The observed heterogeneity in risk could also be caused by environmental factors shared between family members or by differences in risk between mutations. To our knowledge, thus far the only study investigating modifiers of CRC risks for *MUTYH* mutation carriers was on the relationship with hormone replacement therapy, which reported no evidence of interaction between hormone replacement therapy and *MUTYH* mutations.[3]

In this study of 12 variants of *MUTYH* mutations, 93% of the *MUTYH* mutations were Y179C and G396D (Supplementary Table 2); consistent with a previous study of Caucasians.[14] We found CRC risk was higher for monoallelic carriers of Y179C than for G396D; consistent with previous studies.[3, 15] However, given our approach of genotyping for 12 mutations by MS and WAVE followed by confirmatory Sanger sequencing of *MUTYH* in carriers (Materials and Methods), there is the possibility that we missed other pathogenic mutations in *MUTYH* that were not one of the 12 mutations genotyped. Although we identified additional variants from Sanger sequencing, their pathogenicity was considered inconclusive (unclassified variants) and therefore not included in this analysis. Additional *MUTYH* mutations may reside in different ethnic groups however this cohort was predominantly Caucasian.

We used sophisticated statistical techniques to adjust for ascertainment, to account for residual familial aggregation of disease and therefore avoid bias, and to use data for all family members, whether genotyped or not, and therefore maximized statistical power and avoided survival bias.

In conclusion, using the largest international study to date we have produced unbiased estimates of CRC risks for *MUTYH* mutation carriers which are the most precise and reliable currently available. In addition to the confirmed very high risk of CRC to biallelic *MUTYH* mutation carriers, CRC risk for monoallelic mutation carriers depends on family history and can be sufficiently high to warrant consideration of more intensive CRC screening than for the general population.

Figure 1 Cumulative risk of colorectal cancer for (A) male and (B) female *MUTYH* mutation carriers. Note that the risks for a monoallelic carrier with an affected first-degree relative (FDR) who is either untested, a noncarrier or a monoallelic carrier are virtually identical (see Table 1) so the unbroken, darker grey lines cannot be distinguished in the figure.

REFERENCES

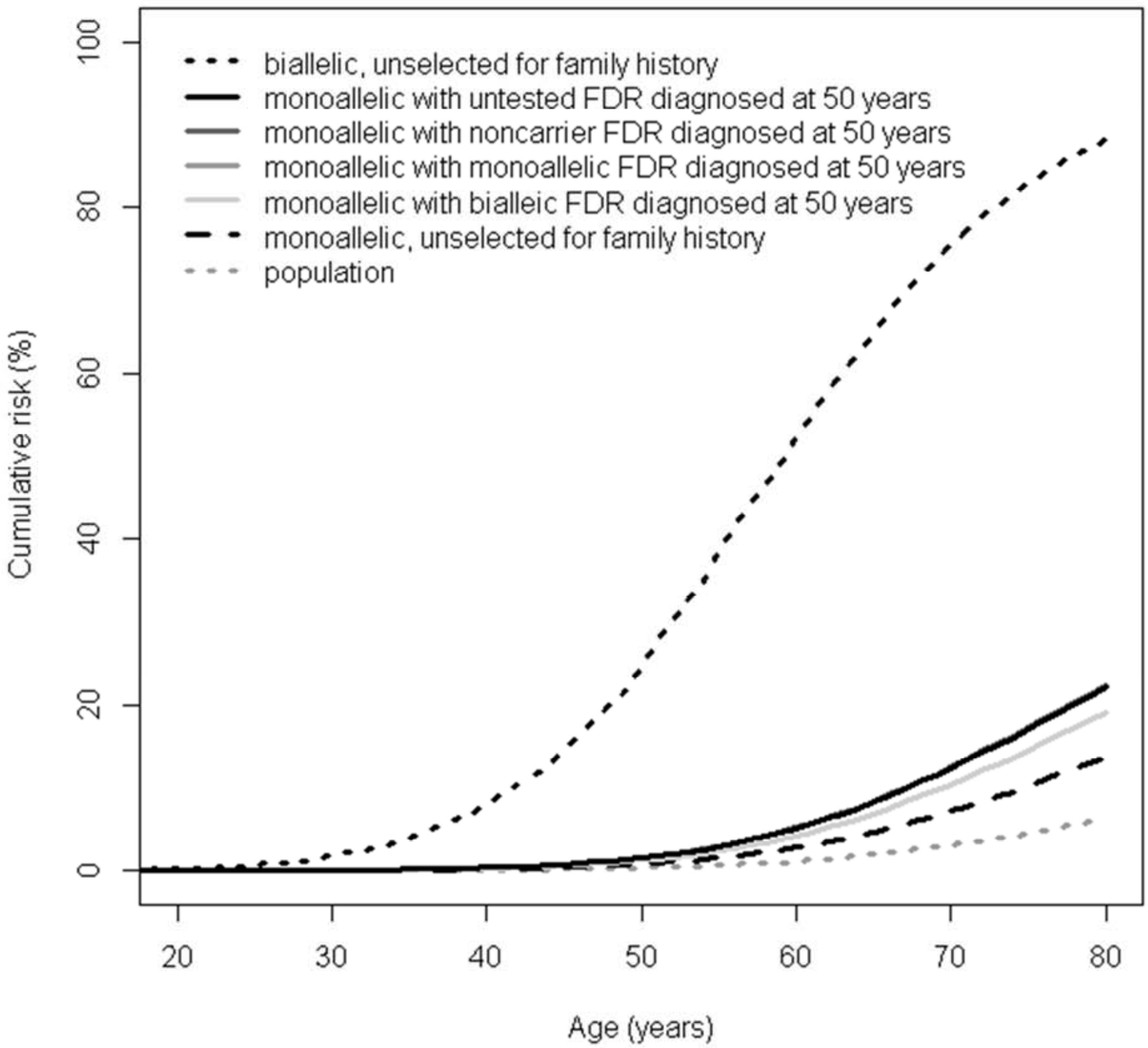
1. Al-Tassan N, et al. Nat Genet 2002;30:227.
2. Cleary SP, et al. Gastroenterology 2009;136:1251-1260.
3. Theodoratou E, et al. Br J Cancer 2010;103:1875-84.
4. Win AK, et al. Fam Cancer 2011;10:1-9.
5. Win AK, et al. Int J Cancer 2011;129:2256-2262.
6. **Grover S, Kastrinos F**, et al. JAMA 2012;308:485-92.
7. Pharoah PDP, et al. Nat Genet 2002;31:33-36.
8. Lubbe SJ, et al. J Clin Oncol 2009;27:3975-3980.
9. **Farrington SM, Tenesa A**, et al. Am J Hum Genet 2005;77:112-119.
10. Nieuwenhuis MH, et al. Gut 2012;61:734-8.
11. Macrae F, et al. Gut 2013;62:657-9.
12. Buecher B, et al. Fam Cancer 2012;11:321-8.
13. Jones N, et al. Gastroenterology 2009;137:489-494.
14. Cheadle JP, Sampson JR. DNA Repair 2007;6:274-279.
15. Balaguer F, et al. Clin Gastroenterol Hepatol 2007;5:379-387.

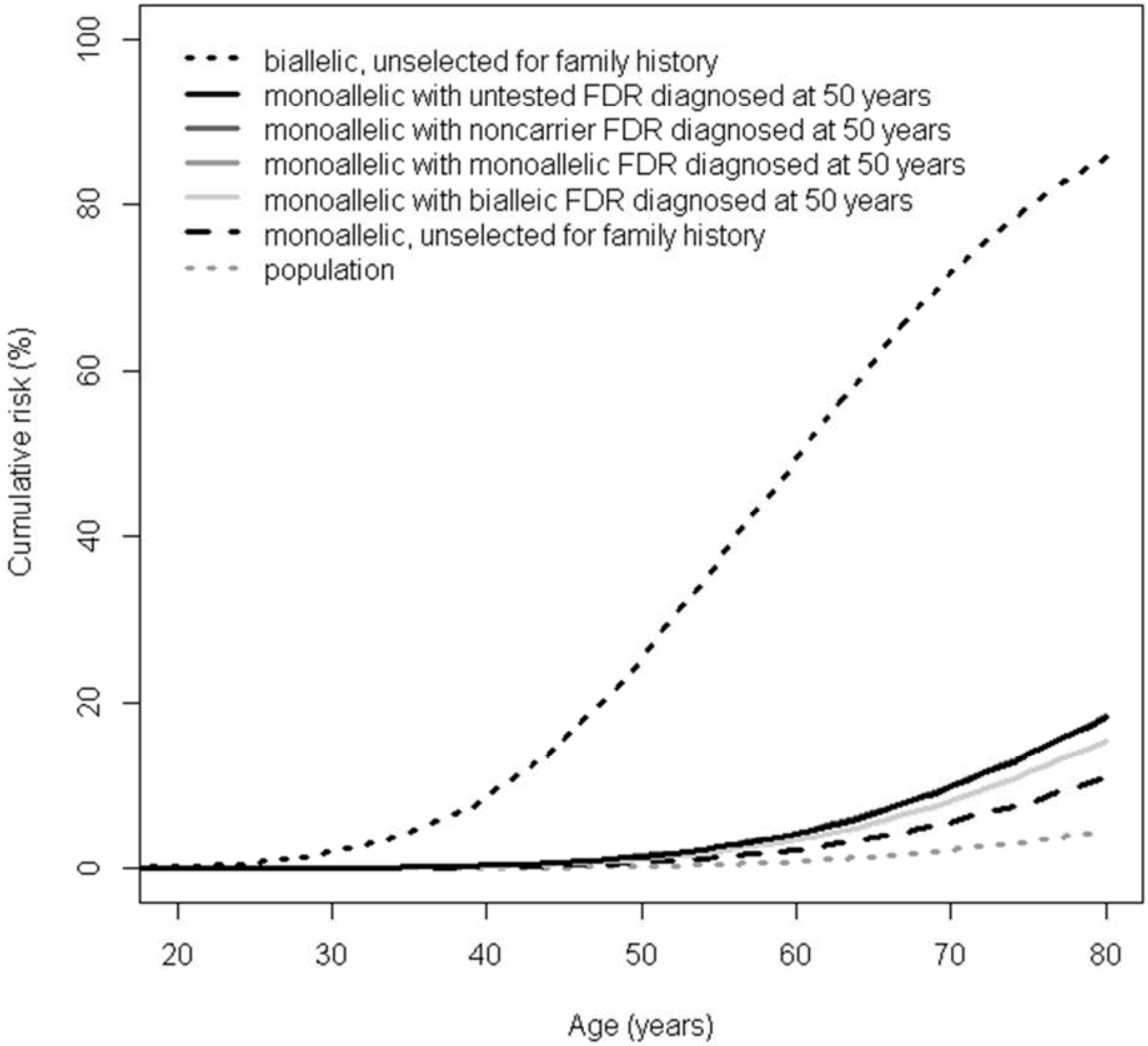
Author names in bold designate shared co-first authors.

Table 1 Cumulative risks (95% confidence intervals) of colorectal cancer for biallelic and monoallelic *MUTYH* mutation carriers

Age (years)	General population	Biallelic mutation carriers irrespective of family history	Monoallelic mutation carriers irrespective of family history	with untested FDR diagnosed at 50 years	with non-carrier FDR diagnosed at 50 years	with monoallelic FDR diagnosed at 50 years	with bialleic FDR diagnosed at 50 years
Male							
30	0.01	1.8 (0.4-6.8)	0 (0-0.1)	0.1 (0.1-0.1)	0.1 (0-0.1)	0.1 (0-0.1)	0.1 (0-0.1)
40	0.07	8.1 (2.1-25.2)	0.2 (0.1-0.3)	0.4 (0.3-0.6)	0.4 (0.2-0.7)	0.4 (0.2-0.6)	0.3 (0.2-0.5)
50	0.3	24.8 (7.7-57.1)	0.8 (0.5-1.3)	1.6 (1.0-2.5)	1.6 (1.0-2.5)	1.6 (1.0-2.5)	1.3 (0.8-1.9)
60	1.1	52.3 (21.8-85.4)	2.8 (1.8-4.5)	5.2 (3.4-7.8)	5.2 (3.4-7.9)	5.2 (3.4-7.7)	4.2 (2.8-6.3)
70	2.9	75.4 (41.2-96.6)	7.2 (4.6-11.3)	12.4 (8.6-17.5)	12.5 (8.6-17.7)	12.4 (8.6-17.4)	10.4 (7.0-15.0)
80	6.2	88.2 (58.4-99.3)	13.6 (8.8-21.1)	22.2 (16-29.8)	22.3 (16.1-30)	22.2 (16.1-29.9)	19.1 (13.2-26.6)
Female							
30	0.01	2 (0.7-5.5)	0 (0-0.1)	0.1 (0.1-0.1)	0.1 (0.1-0.1)	0.1 (0.1-0.1)	0.1 (0-0.1)
40	0.06	8.7 (3.1-20.7)	0.2 (0.1-0.3)	0.4 (0.2-0.6)	0.4 (0.2-0.7)	0.4 (0.2-0.6)	0.3 (0.2-0.5)
50	0.3	25.4 (10.8-49.1)	0.8 (0.5-1.2)	1.5 (0.9-2.3)	1.5 (0.9-2.4)	1.5 (0.9-2.4)	1.2 (0.8-1.8)
60	0.8	49.4 (25.3-76.3)	2.3 (1.5-3.7)	4.3 (2.8-6.4)	4.3 (2.8-6.6)	4.3 (2.8-6.5)	3.5 (2.2-5.2)
70	2.1	71.7 (44.5-92.1)	5.6 (3.6-8.8)	9.9 (6.7-14.2)	10 (6.7-14.4)	9.9 (6.7-14.2)	8.2 (5.4-12.0)
80	4.4	85.7 (61.8-97.8)	10.9 (7-16.8)	18.2 (12.8-24.9)	18.3 (12.9-25.2)	18.2 (12.8-25.0)	15.4 (10.4-21.9)

FDR, first-degree relative; CRC, colorectal cancer.


biallelic, unselected for family history
monoallelic with untested FDR diagnosed at 50 years
monoallelic with noncarrier FDR diagnosed at 50 years
monoallelic with monoallelic FDR diagnosed at 50 years
monoallelic with bialleic FDR diagnosed at 50 years
monoallelic, unselected for family history
population
Cumulative risk (%)
0
20
40
60
80
100
Age (years)
20
30
40
50
60
70
80


biallelic, unselected for family history
monoallelic with untested FDR diagnosed at 50 years
monoallelic with noncarrier FDR diagnosed at 50 years
monoallelic with monoallelic FDR diagnosed at 50 years
monoallelic with bialleic FDR diagnosed at 50 years
monoallelic, unselected for family history
population
Cumulative risk (%)
0
20
40
60
80
100
20
30
40
50
60
70
80
Age (years)

MATERIALS AND METHODS

Study Sample

Subjects were from families for which at least one family member had been identified as carrying a germline monoallelic or biallelic mutation in *MUTYH*. These families were recruited by the Colon Cancer Family Registry via probands between 1997 and 2007. Population-based probands were recently diagnosed colorectal cancer cases from state or regional population cancer registries in the USA (Washington, California, Arizona, Minnesota, Colorado, New Hampshire, North Carolina, and Hawaii), Australia (Victoria) and Canada (Ontario). Clinic-based probands were enrolled from multiple-case families referred to family cancer clinics in the USA (Mayo Clinic, Rochester, Minnesota, and Cleveland Clinic, Cleveland, Ohio), Australia (Melbourne, Adelaide, Perth, Brisbane, Sydney) and New Zealand (Auckland). No cases were ascertained because of having polyps alone. All cases with familial adenomatous polyposis were excluded.

Probands were asked for permission to contact their relatives to seek their enrollment in the Cancer Family Registry. For population-based families, first-degree relatives of probands were recruited at all centers and recruitment was extended to more distant relatives at some centres. For clinic-based families, recruitment was based on availability but attempts were made to recruit up to second-degree relatives of affected individuals (detail in Newcomb et al.[1]). Written informed consent was obtained from all study participants, and the study protocol was approved by the institutional research ethics review board at each centre.

Data Collection

Standardized questionnaires were used to collect information on demographics, personal characteristics, personal and family history of cancer, cancer screening history, history of polyps, polypectomy and other surgeries were obtained by questionnaires from all probands and participating relatives. Reported cancer diagnoses and age at diagnosis were confirmed, where possible using pathology reports, medical records, cancer registry reports and/or death certificates. Standardized protocols were used to collect and prepare blood samples and tumour tissues for genetic testing and laboratory analyses.

MUTYH Mutation Testing

We tested all probands for mutations in the *MUTYH* gene, and also relatives of those with mutations who provided a DNA sample. As described in detail by Cleary et al. [2], genomic DNA extracted from each participant was sent to a central testing facility (Analytic Genetics Technology Centre, Toronto, Canada). DNA was screened for 12 previously identified *MUTYH* mutations: Y179C, G396D, Y104X, R274Q, E480X, Q391X, c.1145delC, c.933+3A>C, c.1437_1439delGGA, R241W, c.1228_1229insGG, c.1187-2A→G using the MassArray MALDI-TOF Mass Spectrometry (MS) system (Sequenom, San Diego, CA). Screening for R241W, c.1228_1229insGG and c.1187-2A→G was discontinued when testing of 6000 samples failed to identify any carriers of these three variants. All samples with MS mobility shifts underwent screening of the entire *MUTYH* coding region, promoter, and splice sites regions by denaturing high-performance liquid chromatography (Transgenomic Wave 3500HT System; Transgenomic, Omaha, NE), to confirm the mutation and to identify additional mutations. All MS-detected variants and WAVE mobility shifts were submitted for sequencing for mutation confirmation (ABI PRISM 3130XL Genetic Analyser).

From the Colon Cancer Family Registry, we identified 273 probands who were known to carry germline mutations in *MUTYH.* We excluded 9 probands who were also known to carry pathogenic mismatch repair germline muations in a mismatch repair gene (Lynch syndrome). Of the remaining 264 probands, 41 were biallelic mutation carriers and 223 were monoallelic mutation carriers. The variants of *MUTYH* mutations of the probands are shown in Supplementary Table 3. The average age at diagnosis of colorectal cancer was 47.7 (standard deviation [SD] 10.0) years in probands with biallelic mutations and 52.0 (SD 11.9) years in probands with monoallelic mutations.

Statistical Analysis

The median, range, mean and standard deviation of the age at colorectal cancer diagnosis were calculated using Stata 12.1[3]. Hazard ratios (HRs), i.e. the age- and sex-specific cancer incidence for carriers divided by that for the general population [4], were estimated using modified segregation analysis[5, 6]. Models were fit by maximum likelihood with the statistical package MENDEL version 3.2[7]. Estimates were appropriately adjusted for the clinic- and population-based ascertainment of families using a combination of retrospective likelihood and ascertainment-corrected joint likelihood [8-11], in which each pedigree's data was conditioned on the proband's genotype, cancer status and age of onset (for population-based families) or on the proband's genotype and the affected statuses and ages of onset of all family members at the time the proband was found to be a *MUTYH* mutation carrier (for clinic-based families).

To model any residual familial aggregation of colorectal cancer risk, a mixed model that incorporated an unmeasured polygene in addition to the major gene[8, 12], was used in the modified segregation analyses (see detail methods in a previous report [13]). This mixed model was used since major gene models (which attribute all familial aggregation to the major gene being studied alone) are often biased[10]. The polygenic part of this mixed model, which models the cumulative effect of a large number of biallelic genes that individually have small effects on cancer susceptibility, was implemented as a hypergeometric polygenic model with four loci[8, 12]. Under this model, the number of disease alleles for each person is approximately normally distributed and is correlated within families with correlation coefficients equal to the kinship coefficients[14].

Estimated cumulative risks (penetrance) of cancers to age t years for carriers living in the USA were calculated by from the relevant population incidences $\lambda_0(\tau)$ at age τ years multiplied and the estimated HR θ with the formula:

$$1 - \exp\left(-\int_0^t \theta\lambda_0(\tau)\, d\tau\right)$$

We estimated the total number of carriers in the study using the same method in previous studies[15, 16]; by summing *MUTYH* carrier probabilities for all individuals, as calculated from Mendel's laws of inheritance, the known genetic relationship of each individual to his or her genotyped relatives (but not affected status) and a population allele frequency of 0.0085.[15] These calculations were performed using R 2.15.0[17] and a modified version of Mendel 3.2.[7]

Observation time for each subject started at birth and ended at first diagnosis of colorectal cancer or other cancer, first polypectomy, last follow-up or death, whichever occurred first. Where age at diagnosis of cancer was not reported (n=43; 16% of all cancer cases), we assumed the age of diagnosis to be one year prior to the last known age or, if last known age was not available, the median age at diagnosis of colorectal cancer for the general population obtained from SEER Cancer Statistics Review (1975–2007)[18].

Effect of family history on disease risks for major gene mutation carriers

In the following section, we describe statistical methods used to derive the age-dependent cumulative risks and HRs for a carrier of a major gene mutation who has an affected first-degree relative, under the genetic mixed model used in this paper. The cumulative risks and HRs are given in equations (2) and (3), respectively, and equation (4) gives approximate HRs for rare diseases or for early-onset forms of common diseases.

Consider two individuals who are indexed by $i = 1$ and 2 and let T_i, G_i and H_i be random variables representing (respectively) the age at disease onset, major gene genotype and polygenotype of individual i. We assume the two individuals are first-degree relatives, so that (H_1, H_2) is a bivariate, normally-distributed random variable, with the correlation coefficient of H_1 and H_2 being $\frac{1}{2}$ and each H_i having mean μ and variance σ^2. We assume that mutations in the major gene are rare, so that we can ignore biallelic carriers and we can assume $P(G_2 = 1 | G_1 = 1) = \frac{1}{2}$ where $G_i = 0$ if i is a non-carrier and $G_i = 1$ if i is a carrier, though the derivation below can easily be changed to incorporate common alleles (and general modes of inheritance). We also assume T_1 and T_2 are conditionally independent given all genotypes and polygenotypes, that T_1 only depends on G_1 and H_1 (and similarly for T_2) and that G_1 and G_2 are independent of H_1 and H_2.

Let $\lambda_0(t)$ be the average incidence of disease at age t years for non-carriers (the average being over the polygene) and recall that the hazard function $\lambda(x)$ of any continuous random variable X is defined to be

$$\lambda(x) = -\frac{d}{dx}\log\bigl(1 - P(X \leq x)\bigr). \tag{1}$$

Then, as in the rest of this paper, we assume that the hazard of T_i (i.e. the incidence of disease for individual i) at age t_i years, conditioned on $G_i = g_i$ and $H_i = h_i$, is equal to $\theta_{g_i} r_i \lambda_0(t_i)$, where θ_{g_i} is the HR associated with major gene genotype g_i and $r_i = e^{h_i}$ is the HR associated with polygenotype h_i. Using the shorthand t_i, g_i and h_i for the events $T_i \leq t_i$, $G_i = g_i$ and $H_i = h_i$ (respectively) in all probability statements below, it therefore follows from (1) that

$$P(t_i|g_i, h_i) = 1 - s_i^{\theta_{g_i} r_i},$$

where $s_i = \exp(-\int_0^{t_i} \lambda_0(\tau)\, d\tau)$ is the survival function to age t_i years for non-carriers. We assume the HR θ_0 for non-carriers is 1, so in order that the average incidence of disease for non-carriers at age t years will equal $\lambda_0(t)$, we need each log-normal random variable e^{H_i} to have an expected value of 1, i.e. we need $\mu = -\sigma^2/2$.

We are interested in the cumulative risk of disease for individual 1 by age t_1 given that he or she is a carrier and that his or her first-degree relative, individual 2, was affected by age t_2. In other words, we want to calculate $\mathrm{P}(\mathrm{t}_1|\mathrm{g}_1, \mathrm{t}_2)$ when $g_1 = 1$. But $\mathrm{P}(\mathrm{t}_1|\mathrm{g}_1, \mathrm{t}_2) = \mathrm{P}(\mathrm{t}_1, \mathrm{t}_2|\mathrm{g}_1)/\mathrm{P}(\mathrm{t}_2|\mathrm{g}_1)$ and

$$\begin{aligned}
\mathrm{P}(\mathrm{t}_1, \mathrm{t}_2 | \mathrm{g}_1) &= \sum_{\mathrm{g}_2 \mathrm{h}_1 \mathrm{h}_2} \mathrm{P}(\mathrm{t}_1, \mathrm{t}_2, \mathrm{g}_2, \mathrm{h}_1, \mathrm{h}_2 | \mathrm{g}_1) = \sum_{\mathrm{g}_2 \mathrm{h}_1 \mathrm{h}_2} \mathrm{P}(\mathrm{t}_1, \mathrm{t}_2 | \mathrm{g}_1, \mathrm{g}_2, \mathrm{h}_1, \mathrm{h}_2) P(\mathrm{g}_2, \mathrm{h}_1, \mathrm{h}_2 | \mathrm{g}_1) \\
&= \sum_{\mathrm{g}_2 \mathrm{h}_1 \mathrm{h}_2} \mathrm{P}(\mathrm{t}_1 | \mathrm{g}_1, \mathrm{h}_1) \mathrm{P}(\mathrm{t}_2 | \mathrm{g}_2, \mathrm{h}_2) P(\mathrm{g}_2 | \mathrm{g}_1) P(\mathrm{h}_1, \mathrm{h}_2) \\
&= \frac{1}{2} \sum_{\mathrm{h}_1 \mathrm{h}_2} P(\mathrm{h}_1, \mathrm{h}_2) \left(\left(1 - s_1^{\theta r_1}\right)\left(1 - s_2^{r_2}\right) + \left(1 - s_1^{\theta r_1}\right)\left(1 - s_2^{\theta r_2}\right) \right) \\
&= \frac{1}{2} \mathbb{E}\left[\left(1 - s_1^{\theta R_1}\right)\left(2 - s_2^{R_2} - s_2^{\theta R_2}\right)\right],
\end{aligned}$$

where $R_i = e^{H_i}$, $\mathbb{E}$ is the expectation functional, and $\theta = \theta_1$ is the HR for carriers of major gene mutations compared to non-carriers. Similarly, $\mathrm{P}(\mathrm{t}_2 | \mathrm{g}_1) = \frac{1}{2} \mathbb{E}[2 - s_2^{R_2} - s_2^{\theta R_2}]$, so the cumulative risk is

$$\mathrm{P}(\mathrm{t}_1 | \mathrm{g}_1, t_2) = \frac{\mathbb{E}\left[\left(1 - s_1^{\theta R_1}\right)\left(2 - s_2^{R_2} - s_2^{\theta R_2}\right)\right]}{\mathbb{E}\left[2 - s_2^{R_2} - s_2^{\theta R_2}\right]} = 1 - \frac{\mathbb{E}\left[s_1^{\theta R_1}\left(2 - s_2^{R_2} - s_2^{\theta R_2}\right)\right]}{\mathbb{E}\left[2 - s_2^{R_2} - s_2^{\theta R_2}\right]}. \quad (2)$$

By (1), the hazard $\lambda(\mathrm{t}_1 | \mathrm{g}_1, t_2)$ of T_1 at t_1 given $\mathrm{G}_1 = 1$ and $T_2 \leq t_2$ is therefore

$$\lambda(\mathrm{t}_1|\mathrm{g}_1,t_2) = -\frac{d}{dt_1}\log\bigl(1-\mathrm{P}(\mathrm{t}_1|\mathrm{g}_1,t_2)\bigr) = -\frac{d}{dt_1}\log\mathbb{E}\bigl[s_1^{\theta R_1}\bigl(2-s_2^{R_2}-s_2^{\theta R_2}\bigr)\bigr]$$

$$= -\frac{\mathbb{E}\left[\frac{d}{dt_1}\left(e^{\theta R_1\log s_1}\right)\bigl(2-s_2^{R_2}-s_2^{\theta R_2}\bigr)\right]}{\mathbb{E}\bigl[s_1^{\theta R_1}\bigl(2-s_2^{R_2}-s_2^{\theta R_2}\bigr)\bigr]}$$

$$= \frac{\mathbb{E}\left[-\frac{d}{dt_1}(\log s_1)\theta R_1 s_1^{\theta R_1}\bigl(2-s_2^{R_2}-s_2^{\theta R_2}\bigr)\right]}{\mathbb{E}\bigl[s_1^{\theta R_1}\bigl(2-s_2^{R_2}-s_2^{\theta R_2}\bigr)\bigr]}$$

$$= \theta\,\lambda_0(\mathrm{t}_1)\,\frac{\mathbb{E}\bigl[R_1 s_1^{\theta R_1}\bigl(2-s_2^{R_2}-s_2^{\theta R_2}\bigr)\bigr]}{\mathbb{E}\bigl[s_1^{\theta R_1}\bigl(2-s_2^{R_2}-s_2^{\theta R_2}\bigr)\bigr]}.$$

So the hazard ratio $\mathrm{HR}(\mathrm{t}_1|\mathrm{g}_1,t_2)$ for T_1 given $\mathrm{G}_1 = 1$ and $T_2 \leq t_2$ (i.e. the relative risk of disease at age t_1 for carriers with an affected first-degree relative) is

$$\mathrm{HR}(\mathrm{t}_1|\mathrm{g}_1,t_2) = \theta\,\frac{\mathbb{E}\bigl[R_1 s_1^{\theta R_1}\bigl(2-s_2^{R_2}-s_2^{\theta R_2}\bigr)\bigr]}{\mathbb{E}\bigl[s_1^{\theta R_1}\bigl(2-s_2^{R_2}-s_2^{\theta R_2}\bigr)\bigr]}. \qquad (3)$$

It is unlikely that a simple formula can be given for these expectations because, for instance, $\mathbb{E}\bigl[s_1^{\theta R_1}\bigr] = M_{R_1}(\theta\log s_1)$ where M_{R_1} is the moment-generating function of the log-normal random variable R_1, and no closed-form expression for M_{R_1} is known. However, the expectations can be readily evaluated by simulating draws of (H_1, H_2) from the multivariate normal distribution described above.

We can also derive approximations to these expectations if we restrict our attention to rare diseases or to early-onset forms of common diseases. For in these cases, each $s_i \approx 1$ so $|\log s_i| \ll 1$ and we can evaluate approximations to the above expectations to a first order in $\log s_i$. Using $s_i^{\theta R_i} = e^{\theta R_i\log s_i} \approx 1 + \theta R_i\log s_i$ we have

$$\mathbb{E}\left[s_1^{\theta R_1}\left(2 - s_2^{R_2} - s_2^{\theta R_2}\right)\right]$$

$$\approx \mathbb{E}[2(1 + \theta R_1 \log s_1) - (1 + \theta R_1 \log s_1$$

$$+ R_2 \log s_2) - (1 + \theta R_1 \log s_1 + \theta R_2 \log s_2)] = -(\theta + 1) \log s_2\, \mathbb{E}[R_2]$$

$$= -(\theta + 1) \log s_2.$$

Simlarly, $\mathbb{E}\left[R_1 s_1^{\theta R_1}\left(2 - s_2^{R_2} - s_2^{\theta R_2}\right)\right] \approx -(\theta + 1) \log s_2\, \mathbb{E}[R_1 R_2]$. But $R_1 R_2 = e^{H_1 + H_2}$ is log-normally distributed and $H_1 + H_2$ has mean $2\mu = -\sigma^2$ and variance $3\sigma^2$, so $\mathbb{E}[R_1 R_2] = e^{\sigma^2/2}$. Therefore

$$\mathrm{HR}(\mathrm{t}_1 | \mathrm{g}_1, t_2) \approx \theta e^{\sigma^2/2} \tag{4}$$

to first order in $\log s_1$ and $\log s_2$. Setting θ, the major gene HR, equal to 1 then gives the formula derived in the appendix of Pharoah et al.[19]

REFERENCES

1. Newcomb PA, et al. Cancer Epidemiol Biomarkers Prev 2007;16:2331-2343.

2. Cleary SP, et al. Gastroenterology 2009;136:1251-1260.

3. StataCorp. Stata Statistical Software: Release 12. College Station, TX: StataCorp LP, 2011.

4. Parkin DM, et al. Cancer Incidence in Five Continents, Vol. VII (1988-1992). Lyon: IARC CancerBase No.7, 2005.

5. Antoniou AC, et al. Genet Epidemiol 2001;21:1-18.

6. Lange K. Mathematical and statistical methods for genetic analysis. New York: Springer, 2002.

7. Lange K, et al. Genet Epidemiol 1988;5:471-472.

8. Antoniou AC, et al. Genet Epidemiol 2001;21:1-18.

9. Quehenberger F, et al. J Med Genet 2005;42:491-496.

10. Gong G, et al. Genet Epidemiol 2010;34:373-381.

11. Kraft P, et al. Am J Hum Genet 2000;66:1119-31.

12. Cannings C, et al. Adv Appl Prob 1978;10:26-61.

13. Dowty JG, et al. Hum Mutat 2013;34:490-7.

14. Lange K. Mathematical and statistical methods for genetic analysis: Springer, 2002.

15. Win AK, et al. Int J Cancer 2011;129:2256-2262.

16. Baglietto L, et al. J Natl Cancer Inst 2010;102:193-201.

17. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2008.

18. Ries L, et al. SEER Cancer Statistics Review, 1975-2000. Bethesda, MD: National Cancer Institute, 2003.

19. Pharoah PDP, et al. Nat Genet 2002;31:33-36.

SUPPLEMENTARY MATERIALS

Supplementary 1: Materials and Methods

Supplementary 2: Supplementary Table 1

Hazard ratios (95% confidence intervals) of colorectal cancer for biallelic and monoallelic *MUTYH* mutation carriers

Supplementary 3: Supplementary Table 2

Variants of *MUTYH* mutations in probands

Supplementary 4: Supplementary Figure 1

Cumulative risk of colorectal cancer (CRC) to age 70y for counselees in six different scenerios. We assumed that CRC to be diagnosed at or before age 70y for affected persons, and last known age to be age 70y for unaffected persons.

Legend: +/+, biallelic *MUTYH* mutation carrier; +/-, monoallelic *MUTYH* mutation carrier; -/-, no *MUTYH* mutation; ?, ungenotyped for *MUTYH*; filled in symbol, CRC; <70, age of CRC diagnosis (years); arrow, counselee.

Supplementary 5: Supplementary Figure 2

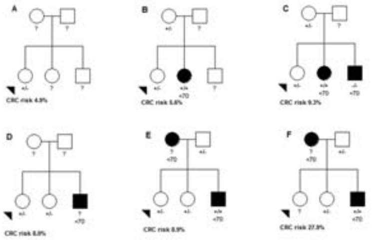
Age-specific hazard ratios for monoallelic mutation carriers with affected first-degree relatives (FDRs) diagnosed by certain ages (grey lines) as well as the HRs given in equation 4 of the Materials and Methods and derived under an early-onset approximation (black lines).

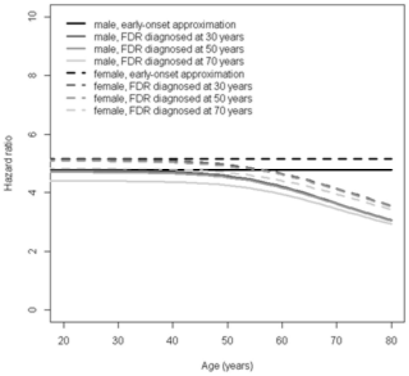
Supplementary Table 1 Hazard ratios (95% confidence intervals) of colorectal cancer for biallelic and monoallelic *MUTYH* mutation carriers

		Biallelic mutation carriers	Monoallelic mutation carriers
Sex			
	Male	108 (25.9–454)	2.46 (1.54–3.93)
	Female	129 (43.7–380)	2.67 (1.67–4.26)
Variants			
	Y179C	115 (23.3–569)	4.81 (3.00–7.71)
	G396D	94.2 (29.2–304)	2.42 (1.48–3.98)

Supplementary Table 2 Variants of *MUTYH* mutations in probands

Variant 1	Variant 2	No of probands
Biallelic carriers		**41**
1187G>A, GGT>GAT, G396D	1187G>A, GGT>GAT, G396D	14
536A>G, TAC>TGC, Y179C	536A>G, TAC>TGC, Y179C	6
1187G>A, GGT>GAT, G396D	536A>G, TAC>TGC, Y179C	10
1187G>A, GGT>GAT, G396D	821G>A, CGG>CAG, R274Q	1
1187G>A, GGT>GAT, G396D	933+3A>C, splicesite	2
536A>G, TAC>TGC, Y179C	933+3A>C, splicesite	1
536A>G, TAC>TGC, Y179C	734G>A, CGT>CAT, R245H	2
536A>G, TAC>TGC, Y179C	312C>A, TAC>TAA, Y104X	1
536A>G, TAC>TGC, Y179C	1437_1439delGGA, in-frame deletion	1
536A>G, TAC>TGC, Y179C	1145delC	2
1171C>T, CAG>TAG, Q391X	1437_1439delGGA, in-frame deletion	1
Monoallelic carriers		**223**
1187G>A, GGT>GAT, G396D		150
536A>G, TAC>TGC, Y179C		55
821G>A, CGG>CAG, R274Q		13
1437_1439delGGA, in-frame deletion		2
1145delC		2
933+3A>C, splicesite		1


A
?
?
+/-
?
?
CRC risk 4.9%
B
+/-
?
+/-
+/+
<70
?
CRC risk 5.6%
C
+/-
?
+/-
+/+
<70
-/-
<70
CRC risk 9.3%
D
?
?
+/-
+/-
?
<70
CRC risk 8.0%
E
?
<70
+/-
+/-
+/-
+/+
<70
CRC risk 8.9%
F
?
<70
+/-
?
+/-
+/+
<70
CRC risk 27.9%


male, early-onset approximation
male, FDR diagnosed at 30 years
male, FDR diagnosed at 50 years
male, FDR diagnosed at 70 years
female, early-onset approximation
female, FDR diagnosed at 30 years
female, FDR diagnosed at 50 years
female, FDR diagnosed at 70 years
Hazard ratio
0
2
4
6
8
10
20
30
40
50
60
70
80
Age (years)