UNIVERSITAT AUTONOMA DE BARCELONA

DEVELOPMENT AND VALIDATION OF A PREDICTIVE MODEL FOR IDENTIFICATION OF *MLH1* AND *MSH2* MUTATION CARRIERS IN LYNCH SYNDROME

Judith Balmaña Gelpí

Doctoral Thesis

Directors Joan Brunet

José Baselga

Barcelona, 2009

The current thesis has been supervised by Sapna Syngal, M.D., M.P.H.

"It's tough making predictions, especially about the future"

Attributed to many individuals, including Yogi Berra. But, as Yogi himself said "I really didn't say everything I said".

"Make everything as simple as possible, but not simpler".

Albert Einstein.

Acknowledgements/Agraïments:

To Sapna Syngal, who mentored me during my stay at Dana-Farber Cancer Institute. Her leadership, encouragement and generosity have guided me during the last years in my academic development and the submission of this thesis.

To all my colleagues at Dana-Farber Cancer Institute, Dave, Elena, Fay, Rowena, and at the Erasmus Medical Center, Ewout, with whom I have shared unforgettable moments of work and joy.

Als meus directors de tesi, Joan Brunet i José Baselga, per la seva promoció constant i recolzament per dur a terme aquesta tesi doctoral.

En Francesc Balaguer i Antoni Castells, com a representants del grup d'investigadors de l'EPICOLON, per compartir amb mi el seu entusiasme en la recerca del càncer colorrectal hereditari.

A Carmen Alonso, per mostrar-me amb el seu exemple la riquesa humana de la nostra feina.

A totes aquelles persones i companys amb qui he compartit moments intensos de feina i amb qui he disfrutat treballant, Claudi, Judit, Ester, Cristina, Eduardo, Meritxell, Javier, Neus, Dani, i molts més.

Als meus pares, les meves germanes i germà, que em fan costat de manera incondicional.

A l'Alberto, la Jana i la Claudia, de qui aprenc dia a dia.

INDEX

1. **Background** 7
    - **1.1. History of Lynch syndrome** 7
    - **1.2. Prevalence** 11
    - **1.3. Molecular basis** 11
    - **1.4. Clinical features** 13
    - **1.5. Screening and follow-up recommendations** 15
    - **1.6. Identification of individuals at risk** 17
    - **1.7. Development of predictive models** 17
2. **Goals** 24
3. **Articles** 25
    - **3.1.1.** Article 1. *Prediction of MLH1 and MSH2 mutations in Lynch syndrome*
    - **3.1.2.** Article 2. *Comparison of predictive models, clinical criteria and molecular tumour screening for the identification of patients with Lynch syndrome in a population-based cohort of colorectal cancer patients*
4. **Discussion** 26
5. **Conclusions** 34
6. **Next steps** 35
7. **Appendix** 36
    - **7.1. Appendix 1.** *Validation and extension of the PREMM$_{1,2}$ Model in a population-based cohort of colorectal cancer patients*
    - **7.2. Appendix 2.** *Data reduction for prediction: A case study on robust coding of age and family history for the risk of having a genetic mutation*

7.3. Appendix 3. *Phenotype comparison of MLH1 and MSH2 mutation carriers in a cohort of 1,941 individuals undergoing clinical genetic testing in the United States.*

7.4. Appendix 4. *Mismatch repair gene analysis in Catalonian families with colorectal cancer*

7.5. Appendix 5. *What is the hereditary non-polyposis colorectal cancer syndrome?*

8. References.. 41

1. Background

1.1. History of Lynch syndrome

Hereditary non-polyposis colorectal cancer (HNPCC) is an autosomal-dominant disorder with a marked increase in cancer susceptibility, especially cancer of the colorectum and the endometrium. It is characterized by early appearance of tumors, predominantly localized in the proximal colon and occasionally multiple. The disease was described for the first time in 1913 by Dr. Aldred Warthin. In 1895 he noticed that his seamstress was very depressed and questioned her about her grief. She told him that she would die of cancer because most of her relatives had died of colon, gastric or uterine cancer. In fact, she did die at a young age of endometrial carcinoma. He studied her family, now known as "Warthin's Family G"[1] with documentation about her pedigree chart showing the genealogy and pathology for many relatives. This information was not fully appreciated until two extended kindreds were described under the name of Cancer Family Syndrome (CFS) in the mid 1960s[2] by Henry T. Lynch. These two CFS families were named Family "N" (Nebraska) and family "M" (Michigan) and were of particular interest because of the wide anatomic distribution of cancer sites, the large number of individuals with multiple primary cancers, the high incidence of endometrial carcinoma and the multigenerational transmission of cancer with an autosomal dominant inheritance pattern. In 1971 Lynch published an updated review of cancer "Warthin's Family G"[1], which had its roots in Germany. The review was of particular interest because it demonstrated a preponderance of adenocarcinomas of the colon, endometrium and stomach. Interestingly, gastric carcinoma was the predominant cancer in the early generations of the family but was later replaced by colon carcinoma. This change paralleled the decline of gastric carcinoma in the general population and implies an environmental influence in these hereditary cancers. During

the 1970s and early 1980s geneticists and epidemiologists suggested that those families being labeled as CFS were chance clusters of cancer or FAP with inadequate pathologic study. However, by the mid-1980s pedigrees consistent with the CFS began to appear in the literature from around the world[3-5]. The name was changed to HNPCC to clarify the lack of multiple colonic polyps and separate it from the polyposis syndromes. In 1984, Boland and Troncale[6] suggested the term Lynch Syndrome I to describe families with predominantly colon malignancies at an early age and Lynch Syndrome II to describe "cancer families" with colonic and extracolononic malignancies (especially those of the female genital tract). In 1989 the International Collaborative Group on HNPCC (ICG-HNPCC) was established with the aim of developing uniform diagnostic criteria for HNPCC, to improve patient and physician education about the disorder, to establish international collaborative studies and to promote the development of national HNPCC registries. In 1991, the ICG-HNPCC published the so-called "Amsterdam criteria" for the clinical diagnosis of HNPCC[7]. These criteria are set out in Table 1. At the ninth and tenth meeting of the ICG-HNPCC new selection requisites for collaborative studies were proposed that included extracolonic cancers associated with HNPCC. Evaluation of current literature on the tumor spectrum of HNPCC indicated that cancer of the endometrium, stomach, ovaries, small bowel, ureter, renal pelvis, brain, and hepatobiliary tract are all associated with HNPCC[8-10]. Among these tumors, cancer of the endometrium, ureter, renal pelvis, and small bowel have the highest relative risk, and are therefore the most specific for HNPCC. The set of new clinical criteria shown in Table 2 was proposed and accepted by the ICG-HNPCC in 1998. These are the so-called "Amsterdam Criteria II".

Table 1. ICG-HNPCC criteria for the diagnosis of HNPCC (Amsterdam Criteria I).

1. Three or more relatives with colorectal cancer, one of whom must be a first-degree relative of the oher two.
2. At least two generations affected.
3. At least one affected individual to be aged <50 years.
4. Familial adenomatous polyposis excluded.
5. Tumors should be verified by pathological examination.

Table 2. Revised ICG-HNPCC criteria (Amsterdam Criteria II).

1. There should be at least three relatives with an HNPCC-associated cancer (CRC, cancer of the endometrium, small bowell, ureter, or renal pelvis).
2. One should be a first-degree relative of the other two.
3. At least two successive generations should be affected.
4. At least one should be diagnosed before age 50.
5. Familial adenomatous polyposis should be excluded in the CRC case(s) if any.
6. Tumors should be verified by pathological examination.

In 1997, the Bethesda guidelines[11] were developed to identify individuals with colorectal cancer who should be tested for microsatelite instability (MSI), a typical molecular feature of HNPCC tumors[12]. These guidelines were revised in 2004 (table 3). The appropriateness of the name HNPCC was discussed again at an international workshop in Bethesda in 2004[13]. Most participants considered the name inadequate because they realized that HNPCC is a misnomer that that describes a syndrome that, in women, can lead to a predisposition for endometrial cancer. Therefore, the name Lynch syndrome was adapted and is currently being used.

Table 3. Revised Bethesda guidelines[13]

Tumors should be tested for MSI when 1 or more of the following exist:
1. Colorectal cancer diagnosed in a patient who is younger than 50 years. 2. Presence of colorectal cancers that synchronous or metachronous or other tumors associated with Lynch syndrome[#], regardless of age. 3. Colorectal cancer with the MSI-H histology[§] diagnosed in a patient who is less than 60 years of age[¥]. 4. Colorectal cancer diagnosed in one or more first-degree relatives with a Lynch-related tumor, with one of the cancers being diagnosed under age 50 years. 5. Colorectal cancer diagnosed in two or more first- or second-degree relatives with Lynch-related tumors, regardless of age. [#] *Lynch syndrome-related tumors include colorectal, endometrial, stomach, ovarian, pancreas, ureter and renal pelvis, biliary tract, and brain (usually glioblastoma as seen in Turcot syndrome) tumors, sebaceous gland adenomas and keratoacanthomas in Muir-Torre syndrome, and carcinoma of the small bowel.* [§] *Presence of tumor infiltrating lymphocytes, Crohn's-like lymphocytic reaction, mucinous/signet-ring differentiation, or medullary growth pattern.* [¥] *There was no consensus among the Workshop participants on whether to include the age criteria in guideline 3 above; participants voted to keep less than 60 years of age in the guidelines.*

1.2. Prevalence

Lynch syndrome is the most frequent hereditary colorectal cancer syndrome. It is estimated to account for 1-3% of all colorectal cancer patients and endometrial cancer patients[14]. Many studies have been published to attempt to address the prevalence of this syndrome and the heterogeneity is huge, mostly because population-based studies are few, geographical distribution of carriers may differ due to founder mutations, or studies may include testing of different type of mismatch repair (MMR) genes[15-28]. Nevertheless, it is important to highlight that definition of the syndrome has not always been homogeneous and it has hampered its recognition and diagnosis.

In Spain, a population-based study of colorectal cancer patients (EPICOLON study) was performed in 2000-2001 to asses the prevalence of Lynch syndrome in our country[28]. Overall, 1222 patients with newly diagnosed colorectal cancer were included. Of these, 287 (23%) fulfilled the revised Bethesda guidelines and underwent MSI analysis and immunohistochemistry of mismatch repair proteins. Ninety-one patients (7%) had a mismatch repair deficiency. Germline analysis of *MLH1* and *MSH2* genes identified 8 pathogenic mutations (0.7%) in *MSH2* (5 cases) or *MLH1* (3 cases). The study cohort has been used in the current Thesis to compare the performance of predictive models, clinical criteria and molecular tumor screening for the identification of patients with Lynch syndrome.

1.3. Molecular basis

Lynch syndrome is being transmitted with an autosomal dominant pattern. Genes associated with this syndrome belong to the DNA mismatch repair (MMR) system. To date, six of these genes have been well characterized, namely *MSH2*[29], *MLH1*[30-31], *MSH6*[32], *PMS1*[33], *PMS2*[33] and *MSH3*[34]. These genes encode for proteins

involved in repairing DNA mismatches occurred during replication. Tumors lacking this repair facility have been termed MSI+ (microsatellite instability) and express a mutator phenotype where multiple genetic lesions accumulated, mainly in mononucleotide or dinucleotide repeats (CA) n[12,35]. Microsatelites are short repetitive sequences of DNA distributed throughout the human genome which may have non-pathological length polymorphisms. During DNA synthesis, the primer and template strands in a microsatellite can occasionally dissociated and re-anneal incorrectly. This gives rise to heteroduplex DNA molecules, in which the number of microsatellite repeats units in the template and the newly synthesized strand differ. These heterogeneities are known as insertion-deletion loops and together with base-base mismatches are repaired by the MMR system. When the MMR system fails, insertion-deletion loops and base-base mismatches will remain uncorrected, therefore MSI may appear and lead to accumulation of mutations and repeats located in coding sequences in cancer-associated genes such as TGF-β-RII, IGF-RII and BAX and finally lead to colorectal carcinogenesis. Importantly, MSI is already present in the earliest lesions (adenomas) in Lynch syndrome patients, thus indirectly supporting the notion the MSI is more likely to be a cause than a consequence of genetic instability in cancer cells[36-38]. In Lynch syndrome, there is a germline mutation in one allele of any of the known MMR genes, which increases the likelihood of acquiring a second mutation. Tumorigenesis will develop when, in a given cell, a somatic mutation occurs in the remaining allele, leading to a significant deficiency in the MMR system in the target tissue such as colon or endometrium[38].

The MSI+ phenotype is found in approximately 90% of Lynch syndrome tumors, but also in 7-15% of sporadic CRC[39-41]. This sporadic MSI seems to be associated with hypermethylation of the *MLH1* promoter region and epigenetic silencing of this gene may be the cause of MSI in these tumors[42]. In these cases, somatic detection of mutations in the gene BRAF may help to distinguish between a hereditary or somatic dysfunction of the MLH1 gene[43-45].

On the other hand, mutations in the MMR genes lead to loss of expression of the corresponding protein in the tumor. Therefore, immunohistochemistry (IHC) analysis can be performed to assess with gene is being mutated[46-48]. With the availability of antibodies against MMR gene-associated proteins, immunohistochemical evaluation of tumors has gained popularity as the first step in checking for mutations[25]. Lack of staining of MLH1 indicates either a germline mutation or methylation of the promoter, while lack of staining for MSH2 is a strong indicator for either an *MSH2* or *MSH6* mutation. From a clinical perspective, recognizing the heterodimeric partners MSH2-MSH6 and MLH1-PMS2 helps identify the causal mutations in Lynch syndrome by IHC.

1.4. Clinical features

Lynch syndrome is characterized by early-onset (usually before the age of 45-50) CRC with a predominance in the proximal colon[49-50]. Tumors tend to be poorly differentiated (solid or cribiform pattern) or with signet ring cells, and of the mucinous type[51-55]. Some other typical pathological features are the Crohn's-like lymphoid reaction[55] and, apparently, an increased risk of synchronous villous adenomas[52,56]. The

risk of a synchronous CRC in a patient with HNPCC at the time of diagnosis is approximately three times the risk in the general population (7.4% for *MLH1*, 6.7% for *MSH2*, and 2.4% for general population)[52,55]. The rate of metachronous CRC is also higher than in the general population, with an accumulated risk of 40% over the ten years after surgery[50]. Overall the estimated risk for colorectal cancer in those with MMR gene mutations is 70% by age 70 years, with most studies reporting a mean age of diagnosis of the first cancer in the mid 40s. About two thirds of colorectal cancers are right-sided[56-57].

In the tumor spectrum associated with HNPCC families, endometrial neoplasm is the second most common type[58-61]. The estimated lifetime risk for endometrial adenocarcinoma is 40% to 60%, with the mean age at diagnosis around age 50 years[59,62]. The other extracolonic types of tumors associated with Lynch syndrome are cancers of the ovary, uro-epithelium (transitional-cell carcinoma) and kidneys (renal-cell carcinoma), stomach, biliary tract, as well as tumors of the central nervous system in the known Turcot's syndrome[63-65].

Some genotype-phenotype differences have also been reported. There seems to be a higher prevalence of CRC in *MLH1* mutation carriers than in *MLH2* carriers and a younger age of onset in male carriers[66-69]. Endometrial cancer rates and ages of onset do not significantly differ between *MLH1* and *MSH2* mutation carriers[68,71] but other extracolonic, extraendometrial Lynch-associated tumors seem to be more prevalent in *MSH2* mutation carriers[69-70,72]. In general, the frequency of colorectal cancer is lower in *MSH6* carriers and the median age of onset is higher[71,73-74]. Regarding *PMS2* mutation carriers, with very anecdotic data on the risk of cancer in these mutation carriers, lifetime risk of colorectal and endometrial cancer is estimated to be lower and with an

older age of onset than in *MLH1/MSH2* carriers[75]. No increased risk of other Lynch-associated tumors has been reported in heterozygous *PMS2* carriers.

The presence of cutaneous neoplasms such as keratoacanthomas, sebaceous gland adenomas or adenocarcinmas is named Muir-Torre syndrome[76] and it is a phenotypic variant of Lynch syndrome.

1.5. Screening and follow-up recommendations

Identification of individuals at risk for Lynch syndrome allows a presymptomatic diagnosis and the option of offering an intensive screening or preventive measures. Regarding CRC risk, recommendations have centered on colonoscopy. We have current data showing that colonoscopy at 3-year intervals reduce CRC incidence in 62% compared with individuals not undergoing routine surveillance[77-78]. Furthermore, a reduction in CRC mortality has also been observed in these same individuals. Therefore, individuals with known or suspected mutation in a DNA MMR gene or who are at risk based on a documented mutation in the family should be offered colonoscopy every 1-2 years, starting at age of 20 to 25 years (age 30 years in families with an *MSH6* mutation). Missing precursor adenomas and early cancers has led to increased interest in the use of chromogens such as indigo carmine or methylene blue to improve mucosal contrast[79-80]. Regarding endometrial surveillance, it is unclear if it would improve morbidity and mortality for women with Lynch syndrome when more than 75% of women with Lynch syndrome who develop endometrial cancer present with stage 1 disease and the overall 5-year survival rate is over 80%[81-82]. Still,

screening modalities that have been suggested include transvaginal ultrasound and endometrial sampling[83]. In addition, in views of the high risk of ovarian cancer, prophylactic hysterectomy with salpyngo-oophorectomy might be considered in mutation carriers after completion of family planning[84].

The role of screening for extracolonic malignancy in at-risk members of Lynch syndrome families remains unclear. Neoplasms of the stomach and urinary tract, if present in the family, could be screened by means of upper endoscopy and urine cytology with abdominal ultrasound, although no strategy can currently be firmly recommended due to the lack of good data regarding efficacy. A summary of current recommendations is depicted in table 5.

Table 5. Surveillance protocol in Lynch syndrome

Lower age limit	Examination	Interval (years)
20-25	Colonoscopy	1-2
30-35	Gynecological examination, transvaginal ultrasound, aspiration biopsy	1-2
30-35	Gastroduodenoscopy	1-2
30-35	Abdominal ultrasound, and cytology urine	1-2

1.6. Identification of individuals at risk

Common strategies to identify individuals at risk for Lynch syndrome include fulfillment of clinical criteria such as the Amsterdam criteria or the revised Bethesda guidelines[83] which were all developed by consensus of experts. However, the Amsterdam criteria and some components of the Bethesda Guidelines are quite complex to apply. Moreover, each aspect of personal and family history is weighted equally in these guidelines, which encompass multiple diagnoses across generations and were not

designed to determine the likelihood of carrying a genetic mutation for an individual patient. It has been shown that the Amsterdam criteria have limited sensitivity for the identification of mutation carriers and therefore cannot be used as an exclusive means to identify Lynch syndrome patients[85-86]. The Bethesda guidelines were therefore developed as a broader screening tool to identify patients whose tumors should be tested for MSI. However, some component of the Bethesda guidelines remain complex and studies show that they may miss around 10% of mutation carriers[17,25,41]. In addition, it has been shown that health care professionals refer only a fraction of patients who fulfill the Bethesda guidelines for molecular evaluation[87]. Therefore, the optimal method of identifying Lynch syndrome patients is still debated and influx. Some investigators recommend performing IHC on all colorectal cancer specimens[25], while others have shown that fulfillment of the revised Bethesda guidelines followed by MMR deficiency screening is a more efficient strategy[28].

1.7. Development of predictive models

For all of the above reasons, it was necessary to develop and provide health care professionals with accurate and user-friendly tools to help identify individuals at risk of Lynch syndrome, as well as to help delineate the best strategy to pursue genetic evaluation. In addition, quantification or risk could also help communicate with the patient and the family the best strategy to pursue in case of clinical suspicion of a hereditary cancer risk.

Wijnen et al.[88] were the first to develop a multivariable model for *MLH1* and *MSH2* point mutations. They identified three predictors in 184 unrelated kindreds at high risk of familial CRC: fulfillment of the Amsterdam criteria, mean age at diagnosis of CRC, and presence of endometrial cancer in the family. Predictions using this model

can be obtained free of charge via the genetic counseling package CaGene at http://www3.utsouthwestern.edu.

More recently, a predictive model was developed from a familial cancer clinic population which added five variables to the Amsterdam criteria to improve its ability to predict MMR gene mutations (Amsterdam-plus model): number of CRC and endometrial cancers in the family, number of individuals with 2 or more CRC or endometrial primaries, mean age at diagnosis, and number of individuals with 5 or more adenomas.[89]

Both of these models include the rather complex variables within the Amsterdam criteria, were developed using relatively small populations from dedicated high-risk clinics, which may limit their transportability to other settings, and have not been externally validated. These models also do not consider other important factors, such as Lynch-associated tumors other than colorectal and endometrial cancer, and therefore have not become widely adopted in clinical practice.

At the same time of working in the study 1 of this Thesis, two other group of investigators were working on developing two new predictive models, the MMRpredict[90] and the MMRpro model[91]. Barnetson and colleagues[90] analyzed a population-based cohort of 870 patients with colorectal cancer diagnosed before age 55 years. They developed a two-stage prediction model by multivariable logistic regression to estimate the likelihood of finding a mutation in the *MLH1*, *MSH2* and *MSH6* genes (MMRpredict model). Stage 1 of the model used exclusively clinical variables significantly associated with the presence of a mutation, and includes sex, age presence of multiple primary tumors, family history of colorectal cancer and age at which the colorectal cancer was diagnosed, and presence of endometrial cancer in a first degree relative. Stage 2 of the model incorporates data on MSI and IHC to refine the carrier's

prediction. The model was validated in an independent series of 155 colorectal cancer patients under the age of 45. Discriminatory performance was similar between the original and the validation cohort. However, information about calibration of the model (to indicate that the number of individuals predicted to be carriers matched the numbers of observed carriers) has not been provided and the authors warn that prediction in a population older than 55 years is unknown. The model is freely available in an electronic version at http://www1.hgu.mrc.au.uk/Softdata/MMRpredict.php

Chen and colleagues[91] developed a genetic counseling and risk prediction tool that estimates the probability of carrying a deleterious mutation in *MLH1, MSH2* and *MSH6* genes and the probability of developing colorectal or endometrial cancer (MMRpro model). The MMRpro model is evaluated on the basis of detailed personal and family history of colorectal and endometrial cancer including first and second degree relatives. Variables considered in the model are colorectal and endometrial cancer status and age at diagnosis for the counselee and his/her first and second degree relatives; current age or age at last follow-up, if unaffected individuals; MSI results or IHC staining, if tumor available; and result of previous germline testing of *MLH1*, *MSH2*, or *MSH6*. MMRpro uses a mendelian modelling approach and Bayesian rules to translate estimates of mutation prevalence and penetrance of MMR genes into predicted risks of carrying a mutation or developing a colorectal or endometrial cancer. The model was independently validated in 226 clinic-based families for analysis of discrimination and calibration, as well as comparison with current clinical guidelines. Validation showed an area under the ROC curve of 0.83 (95% CI 0.78-0.88) and a ratio of observed to predicted cases of 94% (95% CI 0.84-1.05), which results in higher accuracy than current clinical criteria. Software for performing MMRpro calculations is available free of charge via either the mendelian risk prediction package BayesMendel

at http://astor.som.jhmi.edu/BayesMendel/ or the CaGene package at http://www4.utsouthwestern.edu. MMRpro is a comprehensive risk prediction tool that mimics an autosomal dominant syndrome and incorporates most of clinical and molecular variables associated with Lynch syndrome. However, extracolonic Lynch-associated tumors other than endometrial are not included. Risk prediction through MMRpro requires drawing a pedigree and being familiar with the mendelian risk prediction package. In any case, it is a time consuming exercise that could hamper its implementation in a general non-specialized clinical practice with time constraints. A summary of the above mentioned predictive models are depicted in table 6.

Table 6. Characteristics of prediction models for Lynch syndrome.

Models	Study population and clinical variables	Molecular and genetic strategy	Statistical analysis	Strengths	Limitations	Software
Wijnen et al[88]	184 non-related kindreds Young age at CRC diagnosis in the family; fulfillment of Ams criteria; presence of endometrial cancer in the family	*MLH1* and *MSH2* point mutations	Multivariate logistic regression	1st predictive model developed for Lynch syndrome; simple equation	Developed in a high risk population; Ams criteria as a predictor; not validated; non-individualized risk prediction; large rearrangement analysis not included	CaGene software http://www4.utsouthwestern.edu
Amsterdam -plus[89]	250 non-related kindreds; validated in 94 kindreds Fulfillment of Ams criteria; # of relatives with CRC and endometrial cancer; # of relatives with multiple CRC or endometrial cancer; mean age at dx of CRC and endometrial cancer; # of relatives with > 5 colonic adenomas	*MLH1*, *MSH2* and *MSH6* point mutations	Multivariate logistic regression	Better accuracy compared to Ams criteria	Complex variables; non-individualized risk prediction; software not available	NA
EIFAG model[92]	External validation in 219 high- risk families	*MLH1* and *MSH2* mutations	Mendelian modeling approach	Specifies prior probability of a mutation in *MLH1* or *MSH2*; offers the option of incorporating MSI data	Time consuming; expertise in running the LINKAGE program needed; only CRC and endometrial cancer considered in predictions	http://statgen.dps.unipi.it

MMR Predict[90]	870 CRC patients < 55 years; validation in 155 CRC patients < 45 years Proband: Age at CRC diagnosis; sex; CRC location; synchronous or metachronous tumor; Family: CRC family history with the youngest relative < 50 years or ≥ 50 years; first degree relatives with endometrial cancer	*MLH1, MSH2* and *MSH6* point mutations; *MLH1* and *MSH2* large rearrangements MSI and IHC	Multivariate logistic regression Stage 1: clinical variables Stage 2: refinement with MSI and IHC data	Population-based cohort; prediction of *MLH1*, *MSH2* and *MSH6* and refinement of prediction with molecular data; individualized risk prediction; provides clinical applicability at different cut offs; software available	Developed and validated in a young onset CRC population; extracolonic Lynch-associated neoplasms other than endometrial not included	http://www1.hgu.mrc.ac.uk/Softdata/MMRpredict.php
MMRpro[91]	External validation in 279 individuals from 226 high-risk kindreds Proband and family: CRC and endometrial status and age at diagnosis; current age or age at last follow-up if unaffected; result of MSI, IHC or previous germline genetic testing	*MLH1, MSH2* and *MSH6* point mutations and large rearrangements	Mendelian modeling approach	Offers previous and postsequencing probability of a mutation in *MLH1*, *MSH2* and *MSH6*; accounts for unaffected relatives and family size; considers molecular data; offers risk prediction of developing cancer in unaffected individuals	Time consuming process with pedigree drawing; extracolonic Lynch-associated neoplasms other than endometrial not included	CaGene software http://www4.utsouthwestern.edu http://astor.som.jhmi.edu/BayesMendel/

In this thesis, I present in form of published articles the results of our work on identifying clinical predictors associated with *MLH1* and *MSH2* mutation carriers that turned into development of a prediction model ($PREMM_{1,2}$); and the comparison of two predictive models ($PREMM_{1,2}$ and MMRpredict) with clinical criteria and molecular tumor screening for the identification of CRC patients with Lynch-syndrome in a Spanish population-based cohort of CRC patients**.** A better recognition of these individuals at high risk of colorectal and endometrial cancer may help clinicians provide appropriate recommendations for them and their family members and reduce the incidence of these malignant neoplasms.

2. Goals

1. The first aim was to identify the clinical predictors of finding an *MLH1* or *MSH2* germline mutation according to the personal and family history of an individual and develop a prediction rule that would generate predictions of mutation risk ($PREMM_{1,2}$ Model, Prediction of Mutations in *MLH1* and *MSH2*).

2. The second aim was to compare the performance characteristics of the $PREMM_{1,2}$ and MMRpredict models with clinical criteria and universal molecular screening in a Spanish population-based cohort of CRC patients. The major outcomes of interest were the ability of each strategy to identify individuals who should undergo tumor evaluation for MMR deficiency, as well as predict *MLH1/MSH2* mutation carriers.

3. Articles

Article 1. "Prediction of MLH1 and MSH2 mutations in Lynch syndrome".

Judith Balmaña, MD, David H. Stockwell, MD, MPH, Ewout H. Steyerberg, PhD, Elena M. Stoffel, Md, MPH, Amie M. Deffenbaugh, BS, Julia E. Reid, MStat, Brian Ward, PhD, Thomas Scholl, PhD, Brant Hendrickson, MS, John Tazelaar, MD, Lynn Anne Burbidge, BS, Sapna Syngal, MD, MPH.

JAMA, September 27, 2006-Vol 296, No 12, pages 1469-1478.

Article 2. "Comparison of predictive models, clinical criteria and molecular tumor screening for the identification of patients with Lynch syndrome in a population-based cohort of colorectal cancer patients".

J. Balmaña, F. Balaguer, S. Castellví-Bel, EW. Steyerberg, M. Andreu, X. Llor, R. Jover, A. Castells, S. Syngal, the Gastrointestinal Oncology Group of the Spanish Gastroenterological Association.

J Med Genet 2008; 45:557-563

Prediction of *MLH1* and *MSH2* Mutations in Lynch Syndrome

Judith Balmaña, MD
David H. Stockwell, MD, MPH
Ewout W. Steyerberg, PhD
Elena M. Stoffel, MD, MPH
Amie M. Deffenbaugh, BS
Julia E. Reid, MStat
Brian Ward, PhD
Thomas Scholl, PhD
Brant Hendrickson, MS
John Tazelaar, MD
Lynn Anne Burbidge, BS
Sapna Syngal, MD, MPH

Context Lynch syndrome is caused primarily by mutations in the mismatch repair genes *MLH1* and *MSH2.*

Objectives To analyze *MLH1/MSH2* mutation prevalence in a large cohort of patients undergoing genetic testing and to develop a clinical model to predict the likelihood of finding a mutation in at-risk patients.

Design, Setting, and Participants Personal and family history were obtained for 1914 unrelated probands who submitted blood samples starting in the year 2000 for full gene sequencing of *MLH1/MSH2*. Genetic analysis was performed using a combination of sequence analysis and Southern blotting. A multivariable model was developed using logistic regression in an initial cohort of 898 individuals and subsequently prospectively validated in 1016 patients. The complex model that we have named $PREMM_{1,2}$ (Prediction of Mutations in *MLH1* and *MSH2*) was developed into a Web-based tool that incorporates personal and family history of cancer and adenomas.

Main Outcome Measure Deleterious mutations in *MLH1/MSH2* genes.

Results Overall, 14.5% of the probands (130/898) carried a pathogenic mutation (*MLH1*, 6.5%; *MSH2*, 8.0%) in the development cohort and 15.3% (155/1016) in the validation cohort, with 42 (27%) of the latter being large rearrangements. Strong predictors of mutations included proband characteristics (presence of colorectal cancer, especially ≥2 separate diagnoses, or endometrial cancer) and family history (especially the number of first-degree relatives with colorectal or endometrial cancer). Age at diagnosis was particularly important for colorectal cancer. The multivariable model discriminated well at external validation, with an area under the receiver operating characteristic curve of 0.80 (95% confidence interval, 0.76-0.84).

Conclusions Personal and family history characteristics can accurately predict the outcome of genetic testing in a large population at risk of Lynch syndrome. The $PREMM_{1,2}$ model provides clinicians with an objective, easy-to-use tool to estimate the likelihood of finding mutations in the *MLH1/MSH2* genes and may guide the strategy for molecular evaluation.

JAMA. 2006;296:1469-1478 www.jama.com

Lynch syndrome (also called hereditary nonpolyposis colorectal cancer) is the most common hereditary colorectal cancer syndrome in Western countries, accounting for 2% to 5% of all colorectal cancers (CRCs).[1,2] Lynch syndrome is associated with underlying mutations in the mismatch repair system,[3,4] most commonly in the *MLH1* and *MSH2* genes.[5] Existing clinical criteria to identify Lynch syndrome families include the Amsterdam Criteria[6] and Bethesda Guidelines,[7] and these have been updated, modified, and revised by authorities in the field.[8,9] However, the Amsterdam Criteria and some components of the Bethesda Guidelines remain complex, and the relative importance of the specific aspects of personal and family history included in these guidelines are unclear. In hereditary breast-ovarian cancer syndrome, multiple models have been developed to predict mutations in the *BRCA1* and *BRCA2* genes,[10,11] and these models are widely implemented by health care professionals as they assess their patients' genetic risk.

Using data from a large cohort of individuals undergoing genetic testing of *MLH1* and *MSH2*, we developed a clinical model, the $PREMM_{1,2}$ model (Prediction of Mutations in *MLH1* and

See also pp 1479, 1507, and 1521.

Author Affiliations: Population Sciences Division, Dana-Farber Cancer Institute (Drs Balmaña, Steyerberg, Stoffel, and Syngal), and Division of Gastroenterology, Brigham and Women's Hospital and Harvard Medical School (Drs Stockwell, Stoffel, and Syngal), Boston, Mass; Center for Medical Decision Making, Department of Public Health, Erasmus Medical Center, Rotterdam, the Netherlands (Dr Steyerberg); and Myriad Genetic Laboratories Inc, Salt Lake City, Utah (Mss Deffenbaugh, Reid, and Burbidge, Drs Ward, Scholl, and Tazelaar, and Mr Hendrickson). Dr Balmaña is currently with the Hospital Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain.
Corresponding Author: Sapna Syngal, MD, MPH, Dana-Farber Cancer Institute,44 Binney St, Boston, MA 02115 (ssyngal@partners.org).

©2006 American Medical Association. All rights reserved.

MSH2) to predict the presence of mutations in the *MLH1* and *MSH2* genes based on personal and family history of individuals. For practical application, we have made it available in a Web-based format, so it can be easily accessible to clinicians evaluating individuals with a personal or family history suggestive of Lynch syndrome.

METHODS

Patients

The original cohort for model development consisted of 1219 consecutive unrelated probands who submitted blood samples for full gene sequencing of *MLH1* and *MSH2* to Myriad Genetic Laboratories Inc, Salt Lake City, Utah, starting in 2000. Testing was ordered by health care professionals (mainly geneticists, oncologists, gastroenterologists, or gynecologists) for individuals with a personal or family history suggestive of Lynch syndrome. Data were obtained from the test order form (completed by the health care professional ordering genetic testing) and included the patient's age, sex, and ancestry as well as specific details about personal and family cancer history. We excluded 278 probands for whom the personal and family cancer history were not available and 43 probands who reported a personal history of a Lynch syndrome–associated diagnosis but not age at diagnosis, leaving 898 probands included in the analysis.

Among these 898 probands, there were 2382 relatives reported. We narrowed this group of relatives to include only those who fulfilled the following criteria: (1) first- or second-degree relatives of the proband; (2) affected with Lynch syndrome–associated cancers (of the stomach, ovaries, urinary tract, small intestine, pancreas, bile ducts, brain [glioblastoma multiforme], or sebaceous glands) or colonic adenomas; (3) on the affected side of the family; and (4) age at diagnosis known. This left a total of 1618 reported relatives in the final cohort.

The validation cohort consisted of 1057 consecutive unrelated probands who submitted blood samples for full gene sequencing and large rearrangement analysis of *MLH1* and *MSH2* genes to the same diagnostic laboratory after August 2004. Personal and family history data were obtained in the same way as described for the development cohort. After excluding 41 individuals who did not meet the aforementioned criteria, the validation cohort included 1016 probands.

The test order form used in both cohorts asks specifically for maternal or paternal origin of each relative. When both sides of the family were affected by Lynch syndrome–associated tumors (which occurred in 3 instances of 1914 kindreds), the family history was carefully reviewed to make an assessment of the lineage most likely to be affected. Ethnicity was classified based on the information provided by the health care professional using prespecified categories on the test order form. These data are included because they are relevant for generalizability of the results and demonstrate the heterogeneity of the study population.

The study was investigator-initiated. Data collection and model development occurred independently: collection of clinical data and molecular analyses occurred at Myriad Genetic Laboratories, and an anonymized data set was provided to Dana-Farber/Harvard Cancer Center investigators for all further data analyses. The statistical analysis was conducted by clinical researchers (J.B. and D.H.S.) and an independent statistician (E.W.S.) not affiliated with Myriad Genetic Laboratories. The study was reviewed and approved by the Dana-Farber/Harvard Cancer Center institutional review board; a waiver of informed consent for study participants was obtained because the analysis was performed on deidentified data, without the need for patient contact.

Laboratory Methods

From each sample of blood, DNA from white blood cells was extracted and purified, amplified by polymerase chain reaction, and directly sequenced in forward and reverse directions. For the *MLH1* gene, approximately 2300 base pairs were sequenced, comprising 19 exons and approximately 560 adjacent noncoding intronic base pairs. For the *MSH2* gene, approximately 2800 base pairs were sequenced, comprising 16 exons and approximately 470 adjacent noncoding intronic base pairs. Chromatographic tracings of each amplicon were analyzed by a proprietary computer-based review followed by visual inspection and confirmation. Genetic variants were detected by comparison with a consensus wild-type sequence constructed for each gene. All potential genetic variants were independently confirmed by repeat polymerase chain reaction amplification and sequencing.

For large rearrangement analysis, aliquots of genomic DNA were digested individually with 3 restriction enzymes or combination of enzymes for *MLH1* analysis and 3 restriction enzymes or combinations of enzymes for *MSH2* analysis. Digested DNA was electrophoresed in an agarose gel, transferred to a membrane, and hybridized with a gene-specific probe labeled with phosphate 32. The probe binds to all fragments containing coding sequences of that gene. Autoradiographs and phosphorimages were produced and analyzed for the presence of novel bands and for fragment dosage, from which it was determined which, if any, exons had been deleted or duplicated. Positive and negative controls were run with each batch. All potential mutations were independently confirmed.

Mutations were classified as deleterious, suspected deleterious, uncertain, favor polymorphism, or polymorphism. All nonsense and frameshift mutations that occurred at or before amino acids 733 and 888 of *MLH1* and *MSH2,* respectively, were considered to be deleterious. In addition, specific missense mutations and noncoding intervening sequence mutations were considered to be deleterious on the basis of data derived from linkage analysis of high-risk families, functional assays,

 ©2006 American Medical Association. All rights reserved.

biochemical evidence, and/or demonstration of abnormal messenger RNA transcript processing. Genetic variants for which the available evidence indicates a likelihood, but not proof, that the mutation is deleterious were classified as "suspected deleterious." Examples include mutations that occur at the conserved locations of splice acceptors and splice donors. Missense mutations, mutations that occurred in intronic regions whose clinical significance has not yet been determined, and nonsense and frameshift mutations that occurred distal to amino acid position 733 of *MLH1* and distal to amino acid position 888 of *MSH2* were considered to be variants of uncertain significance. Genetic variants that are highly unlikely to contribute substantially to cancer risk were considered to be polymorphisms. For the purposes of this study, we classified individuals found to have either deleterious or suspected deleterious mutations as mutation-positive. Those with all other genetic variants and polymorphisms were included in the mutation-negative group.

Statistical Methods

Variables related to the proband were the presence and age at diagnosis of CRC, colonic adenomas, endometrial cancer, and other Lynch syndrome–associated cancers (the latter were considered as 1 group). Variables related to the family history included the number of relatives with CRC, endometrial cancer, and other Lynch syndrome–associated cancers, the relationship to the proband (first- vs second-degree), the minimum age at diagnosis for each cancer in the family, and the presence of a relative with more than 1 Lynch syndrome–associated cancer. Because adenomas were reported in only 5% of the relatives (79/1618), we were concerned that this information was unreliable and we therefore did not analyze the effect of adenomas in relatives. Age was treated as a continuous variable, and the effect of age was analyzed separately for each diagnosis. In probands diagnosed as having the same cancer more than once, the age at diagnosis was defined as the youngest age. Restricted cubic spline functions in logistic regression models were used to explore the possibility that the effect of age at diagnosis was nonlinear.[12] In relatives, the minimum age and mean age for any given diagnosis in the family appeared to have similar effects, so we chose to use minimum age for ease of use in clinical practice.

We used univariate analyses to determine how best to include each element of personal and family history in a single multivariable model. We created a variable for probands with 2 or more separate CRCs since this group was reasonably large and had a high prevalence of mutations. Similarly, for relatives with CRC and endometrial cancer, we created variables indicating both the number of affected relatives (1 vs ≥2) and their relationship to the proband (first- vs second-degree). We included 2 variables for each diagnosis in the multivariable model: an indicator variable for the presence or absence of that diagnosis and a variable relating to the age at diagnosis. Finally, the magnitude of the age effect for each diagnosis is presented in decades rather than years for ease of interpretation.

Prediction Rule

We aimed to create a prediction rule that would be simpler to use than a full multivariable model and would generate more robust predictions of mutation risk. All decisions about model specification were based on the development cohort. First, we critically assessed all variables in the model with a *P* value greater than .20 and eliminated 3 that did not achieve this *P* value: age at diagnosis of other Lynch syndrome cancers in the proband, minimum age at diagnosis of other Lynch syndrome cancers in the relatives, and the presence of a relative with multiple cancers. Second, we combined clinically similar age variables with similar statistical effects. One such composite variable included the effects of

Table 1. Demographics of the Study Population*

Characteristics	Development Cohort (n = 898)	Validation Cohort (n = 1016)
Age at testing, median (range), y	49 (5-85)	50 (17-95)
Sex		
Female	566 (63)	742 (73)
Male	332 (37)	274 (27)
Ancestry		
European	637 (71)	813 (80)
Latin American	36 (4)	31 (3)
African	36 (4)	31 (3)
Asian	27 (3)	20 (2)
Native American	18 (2)	10 (1)
Middle Eastern	9 (1)	10 (1)
Not specified	135 (15)	101 (10)
Ordering health care professional		
Geneticist	251 (28)	264 (26)
Gastroenterologists	153 (17)	142 (14)
Oncologist	323 (36)	416 (41)
Obstetrician/gynecologist	54 (6)	61 (6)
Primary care physician	9 (1)	10 (1)
Other	108 (12)	122 (12)
Country of origin		
United States	880 (98)	1006 (99)
Outside United States	18 (2)	10 (1)
Clinical criteria		
Amsterdam II Criteria	260 (29)	274 (27)
Revised Bethesda Guidelines	512 (57)	538 (53)

*Data are expressed as No. (%) unless otherwise indicated.

©2006 American Medical Association. All rights reserved.

the age at diagnosis of CRC or adenoma in the proband as well as the minimum age at diagnosis of CRC in first- and second-degree relatives, and the other reflected the effects of age at diagnosis of endometrial cancer in the proband and first- and second-degree relatives. Finally, we created summary variables for each cancer diagnosis in relatives, in which second-degree relatives were weighted to have half the effect of first-degree relatives.

The modeling process was internally validated by bootstrap resampling. Two hundred random samples were drawn with replacement; predictive models were developed in each sample, including variable selection, and evaluated in the development cohort.[13,14] For external validation, we assessed the performance of the prediction rule derived from the development cohort in the validation cohort. An updated version of the prediction rule was based on logistic regression coefficients estimated from both cohorts, after testing for differences in effects between the development and validation cohorts by statistical interaction terms ("interaction by cohort").

To test the accuracy of the updated model in predicting *MLH1* or *MSH2* mutations, we categorized predicted probabilities of mutation into 5 prespecified but arbitrary categories: 5% or less, 5.1% to 10%, 10.1% to 20%, 20.1% to 40%, and more than 40%. Sensitivity and specificity were calculated and were plotted in a receiver operating characteristic curve. We also included the sensitivity and specificity for the Amsterdam Criteria and revised Bethesda Guidelines[6-9] and assessed predictions made using the Leiden model for 1086 probands with CRC.[15] Multivariable modeling was performed using SAS version 8 software (SAS Institute Inc, Cary, NC), and internal and external validation were performed using S-Plus version 6 software (Insightful Corp, Seattle, Wash). Discrimination between patients with and without mutations was quantified by the area under the receiver operating characteristic curve (AUC), calculated with 95% confidence intervals (CIs). Calibration was assessed graphically and by the Hosmer-Lemeshow goodness-of-fit statistic.[13]

Table 2. Univariate Analysis of Prevalence of *MLH1* and *MSH2* Mutations in Probands According to Clinical History

Characteristics	Prevalence of Mutations in the Development Cohort: No./Total (n = 898)	*P* Value*	Prevalence of Mutations in the Validation Cohort: No./Total (n = 1016)	*P* Value*
Overall mutation-positive rate in the proband	14.5 (130/898)†		15.3 (155/1016)‡	
Proband's history				
CRC				
0	11 (40/362)	<.001	8 (39/466)	<.001
1	15 (72/496)		19 (98/509)	
≥2	45 (18/40)		44 (18/41)	
Adenoma	15 (21/141)	.88	10 (24/231)	.02
Endometrial cancer	24 (22/91)	.006	30 (45/149)	<.001
Other HNPCC cancer	22 (22/100)	.02	19 (18/93)	.25
Multiple HNPCC cancers	28 (34/122)	<.001	39 (50/129)	<.001
No HNPCC-associated diagnosis	4 (7/159)	<.001	6 (12/210)	<.001
First-degree family history				
CRC, No. of relatives				
0	9 (37/396)	<.001	10 (50/491)	<.001
1	16 (59/363)		17 (62/372)	
≥2	24 (34/139)		28 (43/153)	
Endometrial cancer, No. of relatives				
0	13 (101/804)	<.001	14 (126/902)	<.001
1	28 (24/85)		22 (23/105)	
≥2	56 (5/9)		67 (6/9)	
Other HNPCC cancer, No. of relatives				
0	14 (102/758)	.13	14 (127/872)	.26
1	20 (26/129)		19 (24/128)	
≥2	18 (2/11)		25 (4/16)	
Second-degree family history				
CRC, No. of relatives				
0	15 (81/528)	.68	16 (102/651)	.89
1	13 (32/241)		14 (33/229)	
≥2	13 (17/129)		15 (20/136)	
Endometrial cancer, No. of relatives				
0	14 (125/893)	.21	15 (137/939)	.06
1	24 (10/42)		21 (15/69)	
≥2	17 (1/6)		38 (3/8)	

Abbreviations: CRC, colorectal cancer; HNPCC, hereditary nonpolyposis colorectal cancer.
**P* values represent comparison of each characteristic in probands with and without mutations. All *P* values are 2-sided and were obtained by χ^2 tests.
†All were point mutations in *MLH1* or *MSH2*.
‡One hundred thirteen (73%) were point mutations in *MLH1* or *MSH2* and 42 (27%) were large rearrangements in *MLH1* (n = 7) or *MSH2* (n = 35).

RESULTS

Patient Characteristics

The median ages of individuals undergoing genetic testing were 49 and 50 years in the development and validation cohorts, and 63% and 73% of the 898 and 1016 probands were women, respectively (TABLE 1). Patients were mainly of European ancestry, but other ancestries were also represented, including Latin American, African, Asian,

 ©2006 American Medical Association. All rights reserved.

Native American, and Middle Eastern. Ordering health care professionals were mainly geneticists (28% and 26%, respectively) and oncologists (36% and 41%, respectively), but other specialties (ie, gastroenterologists and gynecologists) were also represented. The majority of tests (98% and 99%) were ordered from within the United States, with tests ordered from all 50 states. Overall, 29% and 27% of patient histories fulfilled the Amsterdam I/II Criteria, and 57% and 53% met one of the revised Bethesda Guidelines, respectively. There were no significant differences between demographic and clinical characteristics in the 2 cohorts.

Univariate Analysis

In the development cohort, 14.5% (130/898) of the study individuals were found to have mutations: 6.5% (58/898) had mutations in *MLH1* and 8.0% (72/898) had mutations in *MSH2*. In the validation cohort, the overall prevalence of mutations was 15.3% (155/1016): 5.3% (54/1016) had mutations in *MLH1* and 9.9% (101/1016) had mutations in *MSH2*. Of the 155 mutations detected in the validation cohort, 113 (73%) were point mutations and 42 were large rearrangements, the majority in *MSH2* (83% [35/42]). Mutations were particularly prevalent among probands with 2 or more separate CRCs (45% and 44%, respectively), endometrial cancer (24% and 30%), other Lynch syndrome–associated cancers (22% and 19%), and multiple diagnoses (28% and 39%) (TABLE 2). The prevalence of mutations in the probands increased with increasing numbers of first-degree relatives with CRC or endometrial cancer. As expected, probands with mutations had a younger mean age at CRC diagnosis than those who did not have mutations, and the age at diagnosis of CRC and endometrial cancer was also younger among the relatives of probands with mutations (TABLE 3). In probands, the age difference was most apparent for CRC and colonic adenomas, and in relatives, it was most apparent for CRC and endometrial cancer.

Multivariable Analysis

In the multivariable model (TABLE 4), the risk of finding a mutation was similarly increased in probands diagnosed as having 1 CRC (odds ratio [OR], 2.2; 95% CI, 1.9-2.5), endometrial cancer (OR, 2.5; 95% CI, 2.1-3.1), or other Lynch syndrome–associated cancers (OR, 2.1; 95% CI, 1.7-2.5). Probands with adenomas also had a significantly increased risk of a mutation, with an OR of 1.8 (95% CI, 1.5-2.2). Probands with metachronous or synchronous CRC had a very high OR at 8.2 (95% CI, 5.6-12.0). Among relatives, both the presence and number of first-degree relatives with CRC and endometrial cancer strongly increased the risk of finding a mutation in the proband. Diagnosis of CRC or endometrial cancer at a younger age was clearly associated with an increased risk of finding a mutation (OR per decade younger at time of diagnosis, 1.5; 95% CI, 1.5-1.5).

Prediction Rule

The multivariable model had an AUC of 0.79 (95% CI, 0.76-0.83) at internal validation. The effects were similar in the development and validation cohorts for most predictors (Table 4). However, effects were significantly larger in the validation cohort for one or multiple CRCs in the proband (interaction by cohort, $P=.002$ and $P=.02$) and for endometrial cancer in the proband (interaction by cohort, $P=.01$) without any obvious reason. When only point mutations were considered, external validation of the model in the validation cohort showed an AUC of 0.79 (95% CI, 0.74-0.83), as previously predicted with bootstrapping. Interestingly, the AUC increased to 0.80 (95% CI, 0.76-0.84) when large rearrangement mutations were accounted for in the validation cohort, reflecting that some patients had previously been misclassified as not having a mutation.

The updated prediction rule was based on the combination of the development and validation cohorts (Table 4). A small effect for cohort was incorporated (OR,

Table 3. Comparison of Age at Diagnosis According to Proband Mutation Status

	Development Data Set (n = 898)			Validation Data Set (n = 1016)		
	Mutation-Positive	Mutation-Negative	P Value*	Mutation-Positive	Mutation-Negative	P Value*
Proband, mean age, y, at diagnosis						
Colorectal cancer	40.4	45.5	<.001	45.6	47.7	.09
Adenoma	40.9	45.5	.09	43.8	44.0	.94
Endometrial cancer	45.7	47.3	.54	46.8	49.2	.21
Other HNPCC cancer	47.9	50.4	.49	46.8	48.4	.63
Relatives, mean youngest age, y, at diagnosis in first-/second-degree relatives						
Colorectal cancer	42.2	51.1	<.001	43.5	52.1	<.001
Endometrial cancer	45.8	51.8	.01	46.3	52.0	.008
Other HNPCC cancer	48.6	52.1	.23	51.2	58.6	.003

Abbreviation: HNPCC, hereditary nonpolyposis colorectal cancer.
*P values represent the statistical significance of the comparison of the mean age at diagnosis in probands with and without mutation, and the mean youngest age at diagnosis in relatives of probands with and without mutation. All P values are 2-sided and were obtained by t tests.

©2006 American Medical Association. All rights reserved.

1.28), reflecting the higher prevalence of mutations due to rearrangement analysis in the validation cohort. The equation with the variables included in this updated prediction rule is presented in the BOX. The Web-based clinical model is shown in FIGURE 1 and is accessible to health care professionals at the Dana-Farber Cancer Institute Web site (http://www.dfci.org/premm).

Upon grouping by predicted likelihood of carrying a mutation, patients in the combined cohort were distributed reasonably evenly across 5 categories of risk, with a predicted risk of mutation of 5% or less for 482, 5.1% to 10% for 540, 10.1% to 20% for 460, 20.1% to 40% for 282, and greater than 40% for 150. The model demonstrated excellent ability to discriminate between risk groups (TABLE 5) with an AUC of 0.80 (FIGURE 2A). Sensitivity and specificity depended on the cutoff used for the predicted risk of mutation. If a low cutoff, such as 5%, was used, many patients would be considered for testing, with a sensitivity of 94% but a specificity of 29%. If a high cutoff, such as 40%, was used, specificity would be much better (92%), but many patients with mutations would be missed (sensitivity of 29%). Fulfillment of the Amsterdam II Criteria had a sensitivity of 63% with a 78% specificity, while the revised Bethesda Guidelines had a 74% sensitivity with a specificity of 48% (Figure 2A). Of the 1914 individuals, 105 and 75 mutation carriers did not fulfill the Amsterdam II Criteria or the revised Bethesda Guidelines, respectively, and therefore would not have been tested if only these criteria had been considered (TABLE 6). Compared with the revised Bethesda Guidelines, a 10% cutoff led to testing of fewer patients (47% vs 55%), while missing fewer mutation carriers (15% vs 26%). A safer cutoff of 5% led to more testing (75%) and a lower miss rate (6%). For 1086 probands with CRC, the Leiden model had an AUC of 0.755 compared with 0.806 for the $PREMM_{1,2}$ model (Figure 2B).

Table 4. Multivariable Analysis of the Development and Validation Cohorts for Estimation of the Final Prediction Model (Combined Cohort)

	Odds Ratio (95% Confidence Interval)		
	Development Cohort (n = 898)	Validation Cohort (n = 1016)	Combined Cohort (n = 1914)
Predictors			
Proband			
Unaffected	1.0	1.0	1.0
1 CRC	2.2 (1.9-2.5)	7.0 (6.0-8.1)	3.8 (3.6-4.1)
≥2 CRC	8.2 (5.6-12)	37 (25-55)	16 (14-20)
Adenoma	1.8 (1.5-2.2)	1.5 (1.2-1.7)	1.5 (1.4-1.6)
Endometrial cancer	2.5 (2.1-3.1)	7.1 (6.1-8.2)	4.2 (3.9-4.6)
Other HNPCC cancer	2.1 (1.7-2.5)	1.4 (1.1-1.8)	1.8 (1.6-2.0)
Family history			
CRC			
No family history of CRC	1.0	1.0	1.0
Presence of CRC in first-/second-degree relatives*	2.3 (2.1-2.5)	3.0 (2.8-3.3)	2.6 (2.5-2.7)
≥2 CRC in first-degree relatives	3.1 (2.6-3.6)	4.2 (3.6-4.8)	3.6 (3.4-3.8)
Endometrial cancer			
No family history of endometrial cancer	1.0	1.0	1.0
Presence of endometrial cancer in first-/second-degree relatives*	2.7 (2.4-3.2)	2.7 (2.3-3.1)	2.6 (2.4-2.8)
≥2 Endometrial cancers in first-degree relatives	6.5 (1.8-24)	26 (6.0-113)	12 (6.3-23)
Other HNPCC cancers			
No family history of other HNPCC cancers	1.0	1.0	1.0
Presence of other HNPCC cancers	1.5 (1.4-1.7)	1.4 (1.4-1.6)	1.5 (1.4-1.6)
Age at diagnosis			
CRC†	1.5 (1.5-1.5)	1.4 (1.4-1.4)	1.4 (1.4-1.4)
Endometrial cancer‡	1.3 (1.2-1.4)	1.4 (1.3-1.4)	1.3 (1.3-1.4)
Model performance			
Discrimination			
Area under the ROC curve apparent§	0.81 (0.78-0.85)	0.81 (0.77-0.85)	0.81 (0.78-0.84)
Area under the ROC curve validated	0.79 (0.76-0.83)‖	0.80 (0.76-0.84)¶	0.80 (0.77-0.83)‖
Calibration, predicted vs observed, %	14 vs 14	13 vs 15¶	15 vs 15

Abbreviations: CRC, colorectal cancer; HNPCC, hereditary nonpolyposis colorectal cancer; ROC, receiver operating characteristic.
*Family history was coded as 0 for no affected first- or second-degree relatives, 1 for 1 affected first-degree relative, and an additional 0.5 for 1 or more affected second-degree relatives, such that the family history could have values of 0, 0.5, 1, or 1.5.
†Age effect was considered per decade younger and was averaged for probands with CRC, probands with adenoma, and youngest ages for CRC diagnoses in affected first- or second-degree relatives.
‡Age effect was considered per decade younger and was averaged for probands with endometrial cancer and youngest ages for endometrial cancer diagnoses in affected first- or second-degree relatives.
§Performance was evaluated on the data to derive the final model.
‖Estimated by internal validation (bootstrap method).
¶Estimated by applying the model from the development data set in the validation data set (independent external validation).

COMMENT

Our study reports the prevalence of *MLH1/MSH2* mutations detected from a large and diverse cohort of probands undergoing genetic testing on the basis of clinical history, largely without prior molecular prescreening. We found an overall prevalence of deleterious point mutations in 14.5% of individuals in a cohort of 898 with gene sequencing alone and an increase in prevalence to 15.3% (155/1016) with the addition of Southern blot analysis, with 27% (42/155) of detected mutations corresponding to large rearrangements.

Previous estimates for prevalence of *MLH1* and *MSH2* mutations, the most common genes associated with Lynch syndrome, have ranged from 0.3% to 88% and depend greatly on the popu-

 ©2006 American Medical Association. All rights reserved.

lation studied, prior selection based on microsatellite instability and/or immunohistochemistry, and the sensitivity of the laboratory techniques used for germline mutation detection.[16-22] Although not truly "population-based," our findings likely closely reflect what one would expect to see among individuals currently undergoing direct clinical genetic testing for Lynch syndrome in the US population at risk of the disease.

Because of the large sample size of our cohorts, we were able to precisely quantitate the relative importance of known clinical parameters in Lynch syndrome. The most significant clinical predictors of finding a mutation according to the proband's history were the presence of 2 or more CRCs (associated with an OR of 16) and according to the family history were the number of first-degree relatives with CRC or endometrial cancer. Age at diagnosis was more important as a factor for CRC than for endometrial or other Lynch syndrome cancers. In the latter cases, the clustering of such tumors with CRC in a kindred was much more important than the age at which they were diagnosed. Although history of adenomas could only be assessed for probands, we observed that they were a significant predictor of mutation status in both the derivation and validation cohorts, although less strong than that of a CRC diagnosis.

Despite the fact that Lynch syndrome is the most common hereditary CRC predisposition syndrome, the identification of at-risk families, the approach to molecular evaluation, and clinical management continue to pose significant challenges for researchers and clinicians.[23] One of the main topics of debate has been how to approach the molecular evaluation of patients and their families. Strategies ranging from using existing diagnostic criteria alone to population-based molecular testing of all colorectal tumors using immunohistochemistry have been proposed.[21,22,24,25]

In hereditary breast-ovarian cancer syndrome, several models have been developed for risk stratification of *BRCA1* and *BRCA2* gene mutations and are widely used in clinical practice to assist in genetic evaluation and counseling.[10,11] The availability of similar models has been more limited for Lynch syndrome. The most widely used diagnostic criteria, the Amsterdam Criteria and Bethesda Guidelines, help researchers and clinicians identify individuals and families at risk of this syndrome but include broad and often complex variables that may encompass multiple diagnoses across generations and are not designed to determine the likelihood of carrying a genetic mutation for an individual patient. Wijnen et al[15] developed a multivariable model to identify predictors of *MLH1* and *MSH2* point mutations in 184 unrelated kindreds referred to high-risk clinics that contained 3 predictors of mutations in *MLH1*/*MSH2*: fulfillment of the Amsterdam Criteria, younger mean age at diagnosis of CRC in the family, and presence of endometrial cancer in the kindred. Recently, a quantitative model was developed from a familial cancer clinic population in the United Kingdom,[26] which added 5 variables to the Amsterdam Criteria to improve its ability to predict mismatch repair gene mutations (number of CRC and endometrial cancers in the family, number of individuals with ≥2 CRC or endometrial primaries, mean age at diagnosis, and number of individuals with ≥5 adenomas). Both of these models include the rather complex variables within the Amsterdam Criteria and were developed using relatively small populations from dedicated high-risk clinics. Our larger and heterogeneous study population allowed the $PREMM_{1,2}$ model to be more detailed, taking into account the age at diagnosis in probands and relatives, the presence of colonic adenomas

Box. Equation for $PREMM_{1,2}$ (Prediction of Mutations in *MLH1* and *MSH2*) Model

Predicted probability of a mutation in *MLH1* or *MSH2* = 1/[1 + exp (–L)], where L= –3.87 + 1.33V1 + 2.78V2 + 1.44V3 + 0.59V4 + 0.41V5 + 0.951V6 + 1.27V7 + 0.964V8 + 2.48V9 + 0.404V10 – (0.358)V11/10 – (0.293)V12/10.

V1 = presence of CRC in the proband; V2 = 2 or more CRC in the proband; V3 = endometrial cancer in the proband; V4 = other HNPCC cancer in the proband; V5 = adenoma in the proband; V6 = 1 for presence of 1 CRC in first-degree relative + 0.5 for presence of CRC in second-degree relatives; V7 = 2 or more first-degree relatives with CRC; V8 = 1 for presence of 1 first-degree relative with endometrial cancer + 0.5 for presence of any second-degree relatives with endometrial cancer; V9 = 2 or more first-degree relatives with endometrial cancer; V10 = first- or second-degree relatives with other HNPCC cancer; V11 = sum of ages at diagnosis of CRC/adenoma; V12 = sum of ages at diagnosis of endometrial cancer.

For each diagnosis, brackets are interpreted as [diagnosis] = 1 if the proband or relatives have had the diagnosis, [diagnosis] = 0 otherwise.

For V11 and V12, ages at diagnosis are calculated as [youngest age at diagnosis in years – 45] if the proband or relatives have had the diagnosis. For V11, we consider the sum of 4 ages at diagnosis: age at diagnosis of CRC in the proband, age at diagnosis of adenoma in the proband, age at diagnosis of CRC in a first-degree relative, age at diagnosis of CRC in a second-degree relative. For V12, we consider the sum of 3 ages at diagnosis of endometrial cancer: age in the proband, age in a first-degree relative, and age in a second-degree relative.

If a proband or relative has had a given diagnosis, but the age at diagnosis is unknown, then the age at diagnosis should be estimated. If no age is entered, the model defaults to age at diagnosis = 45 years.

Abbreviations: CRC, colorectal cancer; HNPCC, hereditary nonpolyposis colorectal cancer (other HNPCC-associated cancers: stomach, ovaries, urinary tract, small intestine, pancreas, bile ducts, brain [glioblastoma multiforme], sebaceous glands).

©2006 American Medical Association. All rights reserved.

in probands, and the different effect of each cancer diagnosis among first- and second-degree relatives. This increased level of detail led to better sensitivity and specificity combinations than achieved with the Amsterdam II Criteria and revised Bethesda Guidelines. In contrast with the model of Wijnen et al,[15] which provides a family estimate, $PREMM_{1,2}$ can be used to generate separate probabilities of carrying a mutation for each individual in a family and may help to determine which family member might be most appropriate for testing. More recently, 2 models have included microsatellite instability or immunohistochemistry data to refine the estimated probability of finding a mutation. The first, a mendelian model for determining *MLH1* and *MSH2* carrier probabilities, is based on published estimates of mutation frequencies and cancer penetrances in both mutation carriers and noncarriers and includes clinical microsatellite data.[27] How this model performs on actual data from clinical practice is not yet known. The second model was developed in a large population-based cohort of early onset (<55 years) CRC patients undergoing genetic testing for DNA mismatch repair genes.[28] Data from microsatellite instability and immunohistochemistry were incorporated to refine carrier prediction at different cutoffs. However, its applicability in CRC patients aged 55 years or older or patients with other Lynch syndrome–associated tumors has not been assessed.

As is shown by its good discriminatory ability, the $PREMM_{1,2}$ model may become an effective tool for mismatch repair gene mutation risk stratification, which will complement the existing molecular diagnostic tools and other Bayesian models currently in development. The $PREMM_{1,2}$ model can be used to give accurate estimates of a priori risk of carrying *MLH1/MSH2* mutations. How these risks are translated into clinical decision making depends on a variety of factors, including the availability of comprehensive genetic testing services (sequencing and large rearrangement analysis), the timelines of testing information for clinical management decisions, insurance coverage for testing, and the availability of tissue for analysis. Based on the risk estimate generated from the model and the above factors, a clinician may choose whether genetic evaluation should be pursued, as well as

Figure 1. $PREMM_{1,2}$ Model as Presented on the Web

Proband Information

("Proband" refers to the individual being evaluated. Ideally, this individual should have a cancer or adenoma diagnosis.)

How many separate colorectal cancers has the proband had? None ○ One ○ Two or more ○

If one, what was the age at diagnosis? ☐ *(if unknown, estimate)*

If two or more, what was the youngest age at diagnosis? ☐ *(if unknown, estimate)*

Has the proband had colonic adenoma(s)? Yes ○ No ○

What was the youngest age at diagnosis? ☐ *(if unknown, estimate)*

Has the proband had endometrial cancer? Yes ○ No ○

What was the youngest age at diagnosis? ☐ *(if unknown, estimate)*

Has the proband had another HNPCC-associated cancer? Yes ○ No ○

(Other HNPCC-associated cancers include ovary, stomach, small intestine, urinary tract/kidney, bile ducts, glioblastoma multiforme, sebaceous gland tumors, and pancreas.)

Relatives' Information—First Degree

(Only from affected side of family)

How many first-degree relatives have had colorectal cancer? None ○ One ○ Two or more ○

If one, what was his/her age at diagnosis? ☐ *(if unknown, estimate)*

If two or more, what was the youngest age at diagnosis? ☐ *(if unknown, estimate)*

How many first-degree relatives have had endometrial cancer? None ○ One ○ Two or more ○

If one, what was her age at diagnosis? ☐ *(if unknown, estimate)*

If two or more, what was the youngest age at diagnosis? ☐ *(if unknown, estimate)*

Have any first-degree relatives had another HNPCC-associated cancer? Yes ○ No ○

Relatives' Information—Second Degree

(Only from affected side of family)

How many second-degree relatives have had colorectal cancer? None ○ One ○ Two or more ○

If one, what was his/her age at diagnosis? ☐ *(if unknown, estimate)*

If two or more, what was the youngest age at diagnosis? ☐ *(if unknown, estimate)*

How many second-degree relatives have had endometrial cancer? None ○ One ○ Two or more ○

If one, what was her age at diagnosis? ☐ *(if unknown, estimate)*

If two or more, what was the youngest age at diagnosis? ☐ *(if unknown, estimate)*

Have any second-degree relatives had another HNPCC-associated cancer? Yes ○ No ○

Probability of *MLH1*, *MSH2* Mutation CALCULATE ▸ ☐

$PREMM_{1,2}$ indicates Prediction of Mutations in *MLH1* and *MSH2* model; HNPCC, hereditary nonpolyposis colorectal cancer. Model is available online at http://www.dfci.org/premm. Adapted with permission from the Dana-Farber Cancer Institute.

©2006 American Medical Association. All rights reserved.

the approach to testing, such as prescreening of a tumor specimen with microsatellite instability or immunohistochemistry vs direct germline analysis. Microsatellite instability results were reported for only 47 probands and, hence, were not included in the model. The $PREMM_{1,2}$ model might well be used in the initial assessment of individuals at risk of Lynch syndrome, before microsatellite instability information is available to the clinician. A health care professional may use the tool to decide whether to refer the patient for further risk assessment and whether to pursue molecular prescreening.

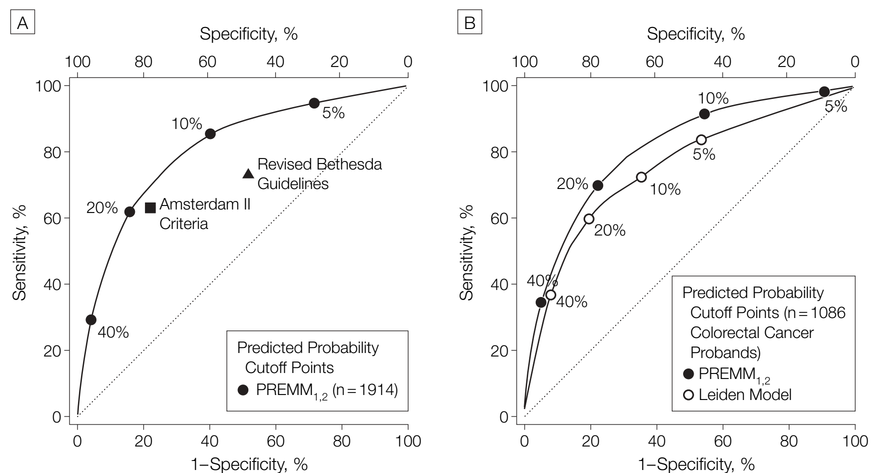
It is important to consider some limitations of our study. The main potential source of error is that our model relies on the clinical history reported by health care professionals on the test order form and the inability to verify diagnoses or collect additional information on certain diagnoses. Previous evidence in the literature shows that accuracy of self-reported family history in first-degree relatives by probands is quite reliable, while it may not be as accurate in second-degree relatives.[29-32] Although reporting errors certainly are likely to occur, the fact that health care professionals are the sources of data likely minimize those based on erroneous diagnoses. Reporting errors are likely to represent both underreporting and overrepresentation of cancer diagnoses. For example, because of time limitations when completing the test order form, health care professionals may only report diagnoses that are considered sufficient to justify ordering the genetic test. Conversely, unaffected relatives are not reported on the test order form and, therefore, overrepresentation of cancer diagnoses may occur in large families in which many unaffected individuals may be present. Despite these inaccuracies, strong predictive effects were found that were similar in the development and validation cohorts, which illustrates the capability of routinely obtained information for selection of patients for further diagnostic workup. We did not have detailed pedigree information on each family and, therefore, could not incorporate the impact of family size and unaffected individuals on the likelihood of carrying a mutation. Finally, our model predicts mutation status only for *MSH2* and *MLH1*. However, we plan to continue to update the model with incorporation of data from *MSH6* sequence analysis when sufficient data are available.

In conclusion, we determined which aspects of personal and family history were most important in predicting the

Table 5. Prevalence of Mutations Observed in *MLH1/MSH2* Among Predicted Model's Risk Groups for Probability of Mutation

Estimated Risk for Probability of a Mutation, %	No. With Mutations/No. Predicted in Risk Group (%)		
	Development Cohort (n = 898)	Validation Cohort (n = 1016)	Combined Cohort (n = 1914)
≤5	7/252 (3)	9/230 (4)	16/482 (3)
5.1-10	15/259 (6)	11/281 (4)	26/540 (5)
10.1-20	35/191 (18)	31/269 (12)	66/460 (14)
20.1-40	39/126 (31)	54/156 (35)	93/282 (33)
>40	34/61 (49)	50/80 (62)	84/150 (56)
Total	**130/898** (14)	**155/1016** (15)	**285/1914** (15)

Figure 2. Sensitivity and Specificity Receiver Operating Characteristic Curves



$PREMM_{1,2}$ indicates prediction of mutations in *MLH1* and *MSH2* model. A, Receiver operating characteristic curve illustrating sensitivity and specificity of the $PREMM_{1,2}$ model at different cutoffs for predicted probabilities. The square represents the sensitivity and 1 − specificity value for fulfillment of the Amsterdam II Criteria and the triangle represents the sensitivity and 1 − specificity value for fulfillment of the revised Bethesda Guidelines at single cutoff points, as these criteria are dichotomous. B, Receiver operating characteristic curve illustrating sensitivity and specificity of the $PREMM_{1,2}$ model and the Leiden model at different cutoffs for predicted probabilities.

Table 6. Comparison of Performance of the $PREMM_{1,2}$ Model With Amsterdam II Criteria and Revised Bethesda Guidelines

Criterion	Sensitivity, %	Specificity, %	No. (%)	
			Individual Candidates for Testing	Mutation Carriers Who Would Not Have Been Tested
$PREMM_{1,2}$ model cutoff value, %				
≥5	94	29	1432 (75)	16 (6)
≥10	85	60	892 (47)	42 (15)
Amsterdam II Criteria	63	78	534 (28)	105 (37)
Revised Bethesda Guidelines	74	48	1050 (55)	75 (26)

Abbreviation: $PREMM_{1,2}$, prediction of mutations in *MLH1* and *MSH2* model.

©2006 American Medical Association. All rights reserved.

outcome of clinical genetic testing in a large, diverse population at risk of Lynch syndrome from across the United States. Our prediction rule includes specific and discrete variables and does not rely on complex combinations of diagnoses across generations. The $PREMM_{1,2}$ model has been externally validated and is available as a user-friendly Web-based model to provide clinicians with an objective tool to estimate the likelihood of finding mutations in the *MLH1* and *MSH2* genes and to help guide the strategy for molecular evaluation.

Author Contributions: Dr Balmaña had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Drs Balmaña and Stockwell contributed equally to this work.
Study concept and design: Balmaña, Stockwell, Deffenbaugh, Syngal.
Acquisition of data: Deffenbaugh, Ward, Scholl, Tazelaar, Burbidge, Syngal.
Analysis and interpretation of data: Balmaña, Stockwell, Steyerberg, Stoffel, Deffenbaugh, Reid, Scholl, Hendrickson, Syngal.
Drafting of the manuscript: Balmaña, Stockwell, Steyerberg, Syngal.
Critical revision of the manuscript for important intellectual content: Balmaña, Stockwell, Steyerberg, Stoffel, Deffenbaugh, Reid, Ward, Scholl, Hendrickson, Tazelaar, Burbidge, Syngal.
Statistical analysis: Balmaña, Stockwell, Steyerberg, Stoffel, Reid.
Obtained funding: Syngal.
Administrative, technical, or material support: Deffenbaugh, Ward, Scholl, Hendrickson, Tazelaar, Burbidge, Syngal.
Study supervision: Balmaña, Deffenbaugh, Scholl, Syngal.

Financial Disclosures: Dr Stoffel reports having received lecture honoraria from Myriad Genetics Laboratories Inc; Dr Syngal reports having received lecture honoraria from and having served on a clinical advisory board for Myriad Genetics Laboratories Inc; and Mss Deffenbaugh, Reid, and Burbidge, Drs Ward, Scholl, and Tazelaar, and Mr Hendrickson were employed by Myriad Genetics Laboratories Inc when the study was conducted. No other disclosures were reported.

Funding/Support: Dr Balmaña was supported by a grant from La Caixa, Barcelona, Spain, Dr Steyerberg by grants from the Trust Fund Erasmus University Rotterdam and a Marx Fellowship from Dana-Farber Cancer Institute, Dr Stoffel from a Junior Faculty Career Development Award from the American College of Gastroenterology, and Dr Syngal by National Institutes of Health/National Cancer Institute grant K24 CA 113433. No grant support was obtained from Myriad Genetics Laboratories Inc to support the study.
Role of the Sponsor: The funding sources played no role in the design and conduct of the study; collection, management, analysis, and interpretation of data; or preparation, review, or approval of the manuscript.
Acknowledgment: We thank Meredith Regan, ScD, Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, for her assistance in data preparation; Jimmy L. Dias, BA, and William B. Dilworth, BA, Interactive Communications, Dana-Farber Cancer Institute, for their technical assistance in the Web-based model development; and Beth M. Ford, BA, Medical Oncology, Dana-Farber Cancer Institute, for her assistance in data collection and administrative issues. None of those listed herein received compensation for their work.

REFERENCES

1. Lynch HT, de la Chapelle A. Genetic susceptibility to non-polyposis colorectal cancer. *J Med Genet*. 1999;36:801-818.
2. Giardiello FM, Brensinger JD, Petersen GM. AGA technical review on hereditary colorectal cancer and genetic testing. *Gastroenterology*. 2001;121:198-213.
3. Leach FS, Nicolaides NC, Papadopoulos N, et al. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell*. 1993;75:1215-1225.
4. Miyaki M, Konishi M, Tanaka K, et al. Germline mutation of *MSH6* as the cause of hereditary nonpolyposis colorectal cancer. *Nat Genet*. 1997;17:271-272.
5. Chung DC, Rustgi AK. The hereditary nonpolyposis colorectal cancer syndrome: genetics and clinical implications. *Ann Intern Med*. 2003;138:560-570.
6. Vasen HF, Mecklin JP, Khan PM, Lynch HT. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). *Dis Colon Rectum*. 1991;34:424-425.
7. Rodriguez-Bigas MA, Boland CR, Hamilton SR, et al. A National Cancer Institute Workshop on hereditary nonpolyposis colorectal cancer syndrome: meeting highlights and Bethesda Guidelines. *J Natl Cancer Inst*. 1997;89:1758-1762.
8. Bellacosa A, Genuardi M, Anti M, Viel A, Ponz de Leon M. Hereditary nonpolyposis colorectal cancer: review of clinical, molecular genetics, and counseling aspects. *Am J Med Genet*. 1996;62:353-364.
9. Umar A, Boland CR, Terdiman JP, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst*. 2004;96:261-268.
10. Couch FJ, DeShano ML, Blackwood MA, et al. *BRCA1* mutations in women attending clinics that evaluate the risk of breast cancer. *N Engl J Med*. 1997;336:1409-1415.
11. Parmigiani G, Berry D, Aguilar O. Determining carrier probabilities for breast cancer-susceptibility genes *BRCA1* and *BRCA2*. *Am J Hum Genet*. 1998;62:145-158.
12. Harrell FE Jr, Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. *J Natl Cancer Inst*. 1988;80:1198-1202.
13. Harrell FE Jr. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.* New York, NY: Springer; 2001.
14. Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54:774-781.
15. Wijnen JT, Vasen HF, Khan PM, et al. Clinical findings with implications for genetic testing in families with clustering of colorectal cancer. *N Engl J Med*. 1998;339:511-518.
16. Aaltonen LA, Salovaara R, Kristo P, et al. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. *N Engl J Med*. 1998;338:1481-1487.
17. Nystrom-Lahti M, Wu Y, Moisio AL, et al. DNA mismatch repair gene mutations in 55 kindreds with verified or putative hereditary non-polyposis colorectal cancer. *Hum Mol Genet*. 1996;5:763-769.
18. Wijnen J, Khan PM, Vasen H, et al. Hereditary nonpolyposis colorectal cancer families not complying with the Amsterdam Criteria show extremely low frequency of mismatch-repair-gene mutations. *Am J Hum Genet*. 1997;61:329-335.
19. Wang Q, Lasset C, Desseigne F, et al. Prevalence of germline mutations of *hMLH1*, *hMSH2*, *hPMS1*, *hPMS2*, and *hMSH6* genes in 75 French kindreds with nonpolyposis colorectal cancer. *Hum Genet*. 1999;105:79-85.
20. Wagner A, Barrows A, Wijnen JT, et al. Molecular analysis of hereditary nonpolyposis colorectal cancer in the United States: high mutation detection rate among clinically selected families and characterization of an American founder genomic deletion of the *MSH2* gene. *Am J Hum Genet*. 2003;72:1088-1100.
21. Pinol V, Castells A, Andreu M, et al. Accuracy of revised Bethesda Guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. *JAMA*. 2005;293:1986-1994.
22. Hampel H, Frankel WL, Martin E, et al. Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). *N Engl J Med*. 2005;352:1851-1860.
23. Terdiman JP. It is time to get serious about diagnosing Lynch syndrome (hereditary nonpolyposis colorectal cancer with defective DNA mismatch repair) in the general population. *Gastroenterology*. 2005;129:741-744.
24. Southey MC, Jenkins MA, Mead L, et al. Use of molecular tumor characteristics to prioritize mismatch repair gene testing in early-onset colorectal cancer. *J Clin Oncol*. 2005;23:6524-6532.
25. Stormorken AT, Bowitz-Lothe IM, Noren T, et al. Immunohistochemistry identifies carriers of mismatch repair gene defects causing hereditary nonpolyposis colorectal cancer. *J Clin Oncol*. 2005;23:4705-4712.
26. Lipton LR, Johnson V, Cummings C, et al. Refining the Amsterdam Criteria and Bethesda Guidelines: testing algorithms for the prediction of mismatch repair mutation status in the familial cancer clinic. *J Clin Oncol*. 2004;22:4934-4943.
27. Marroni F, Pastrello C, Benatti P, et al. A genetic model for determining *MSH2* and *MLH1* carrier probabilities based on family history and tumor microsatellite instability. *Clin Genet*. 2006;69:254-262.
28. Barnetson RA, Tenesa A, Farrington SM, et al. Identification and survival of carriers of mutations in DNA mismatch repair genes in colon cancer. *N Engl J Med*. 2006;354:2751-2763.
29. Murff HJ, Spigel DR, Syngal S. Does this patient have a family history of cancer? an evidence-based analysis of the accuracy of family cancer history. *JAMA*. 2004;292:1480-1489.
30. Aitken J, Bain C, Ward M, Siskind V, MacLennan R. How accurate is self-reported family history of colorectal cancer? *Am J Epidemiol*. 1995;141:863-871.
31. Kerber RA, Slattery ML. Comparison of self-reported and database-linked family history of cancer data in a case-control study. *Am J Epidemiol*. 1997;146:244-248.
32. Mitchell RJ, Brewster D, Campbell H, et al. Accuracy of reporting family history of colorectal cancer. *Gut*. 2004;53:291-295.

 ©2006 American Medical Association. All rights reserved.

Comparison of predictive models, clinical criteria and molecular tumour screening for the identification of patients with Lynch syndrome in a population-based cohort of colorectal cancer patients

J Balmaña,[1] F Balaguer,[2] S Castellví-Bel,[2] E W Steyerberg,[3] M Andreu,[4] X Llor,[5] R Jover,[6] A Castells,[2] S Syngal,[7] for the Gastrointestinal Oncology Group of the Spanish Gastroenterological Association

[1] Department of Medical Oncology, Hospital Vall d'Hebron, Medical Department of Universitat Autònoma de Barcelona, Spain; [2] Department of Gastroenterology, Institut de Malalties Digestives i Metabòliques, Hospital Clínic, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain; [3] Department of Public Health, Erasmus Medical Center-University MC Rotterdam, The Netherlands; [4] Department of Gastroenterology, Hospital del Mar, Barcelona, Spain; [5] Section of Digestive Diseases and Nutrition, University of Illinois at Chicago, Illinois, USA; [6] Department of Gastroenterology, Hospital General Universitario de Alicante, Alicante, Spain; [7] Division of Gastroenterology, Brigham and Women's Hospital, Population Sciences Division, Dana-Farber Cancer Institute, and Harvard Medical School, Boston, Massachusetts, USA

Correspondence to: Dr J Balmaña, Medical Oncology Department, Hospital Universitari Vall d'Hebron. Paseo Vall d'hebron 119-129, Barcelona 08035, Spain; jbalmana@vhebron.net

JB and FB contributed equally to this work.

Presented in part at 43rd Annual Meeting of the American Society of Clinical Oncology, June 2007, Chicago, Illinois, USA.

All authors are listed in the appendix

Received 28 March 2008
Revised 26 May 2008
Accepted 1 June 2008
Published Online First 25 June 2008

ABSTRACT

Background: Several models have recently been developed to predict mismatch repair (MMR) gene mutations. Their comparative performance with clinical criteria or universal molecular screening in a population based colorectal cancer (CRC) cohort has not been assessed.

Methods: All 1222 CRC from the EPICOLON cohort underwent tumour MMR testing with immunohistochemistry and microsatellite instability, and those with MMR deficiency (n = 91) underwent *MLH1/MSH2* germline testing. Sensitivity, specificity and positive predictive value (PPV) of the $PREMM_{1,2}$ and the Barnetson models for identification of *MLH1/MSH2* mutation carriers were evaluated and compared with the revised Bethesda guidelines (RBG), Amsterdam II criteria, and tumour analysis for MMR deficiency. Overall discriminative ability was quantified by the area under the ROC curve (AUC), and calibration was assessed by comparing the average predictions versus the observed prevalence.

Results: Both models had similar AUC (0.93 and 0.92, respectively). Sensitivity of the RBG and a $PREMM_{1,2}$ score $\geqslant$5% was 100% (95% CI 71% to 100%); a Barnetson score >0.5% missed one mutation carrier (sensitivity 87%, 95% CI 51% to 99%). PPVs of all three strategies were 2–3%. Presence of MMR deficiency increased specificity and PPV of predictive scores (97% and 21% for $PREMM_{1,2}$ score $\geqslant$5%, and 98% and 21% for Barnetson $\geqslant$0.5%, respectively).

Conclusions: The $PREMM_{1,2}$ and the Barnetson models offer a quantitative systematic approach to select CRC patients for identification of *MLH1/MSH2* mutation carriers with a similar performance to the RBG.

Lynch syndrome, or hereditary non-polyposis colorectal cancer (HNPCC), is the most common form of inherited colorectal cancer (CRC).[1 2] Individuals with Lynch syndrome are at high risk for developing CRC, as well as tumours in the uterus, ovaries, stomach, small bowel, hepatobiliary system, pancreas, renal pelvis, ureter and bladder.[3] Lynch syndrome is associated with underlying mutations in the DNA mismatch repair (MMR) system,[4–6] most commonly in the *MLH1/MSH2* genes.[7] Mutations in these MMR genes lead to microsatellite instability (MSI) of cancerous tissue[8] and loss of expression of the corresponding protein, which can be tested by immunohistochemistry (IHC). Since intensive cancer screening and prophylactic surgery have been shown to reduce the incidence and mortality of CRC and endometrial cancer in individuals with Lynch syndrome,[9 10] it is important to identify at risk individuals so that clinicians can provide appropriate recommendations for them and their family members.[11]

Initial strategies for identification of individuals at risk of Lynch syndrome were based on the fulfilment of the Amsterdam criteria[12 13] and the Bethesda guidelines (BG).[8 14] However, the Amsterdam I/II criteria have limited sensitivity for the identification of mutation carriers and therefore cannot be used as an exclusive means to identify Lynch syndrome patients.[15 16] The BG were therefore developed as a broader screening tool to identify patients whose tumours should be tested for MSI[8] and have been subsequently revised on several occasions,[2 14] according to empirical data and expert opinion. The optimal method of identifying Lynch syndrome patients is still debated and in flux. Some investigators recommend performing IHC on all CRC specimens,[17] while others have shown that fulfilment of the RBG followed by MMR deficiency screening is a more efficient strategy.[18] However, some components of the BG remain complex and studies show that they may miss around 10% of mutation carriers.[17 19 20] In addition, we have previously shown that medical oncologists refer only a fraction of patients who fulfill the BG for molecular evaluation.[21] Finally, the available clinical criteria only provide a bivariate (yes/no) assessment of risk for Lynch syndrome and do not give a quantitative determination of the likelihood of being a mutation carrier. For all of these reasons, it is necessary to provide health care professionals with accurate and user-friendly tools to help identify individuals at risk of Lynch syndrome, as well as to help delineate the best strategy to pursue genetic evaluation.

In light of these issues, several risk prediction models have recently been developed to estimate the likelihood of a mutation in the MMR genes associated with Lynch syndrome (table 1). The $PREMM_{1,2}$ model was developed to predict mutations in *MLH1/MSH2* in a cohort of individuals at moderate-to-high risk of Lynch syndrome.[15] In contrast, the Barnetson prediction model was developed in a young onset, population based

Table 1 Overview of recent predictive models for mismatch repair (MMR) gene mutation detection in Lynch syndrome

Model	Population	Outcome	Neoplasms included	Method	Predictors in development cohorts
Barnetson *et al*[22]	870 population based colorectal cancer patients <55 years	Point mutations at *MLH1, MSH2, MSH6* and large rearrangements at *MLH1, MSH2*	Colon and endometrial	Two stage model: 1. Multivariate logistic regression analysis 2. Refinement of carrier prediction's performance with MSI and IHC data	Proband: age, gender, tumour's location, multiple tumours Family: CRC age <50 or ⩾50 years, presence of endometrial cancer in 1st degree relatives AUC = 0.82, 95% CI 0.72 to 0.91
$PREMM_{1,2}$[15]	1914 individuals undergoing genetic testing for clinical suspicion of Lynch syndrome	Point mutations and large rearrangements at *MLH1, MSH2*	Colon, endometrial, other Lynch associated tumours, colonic adenomas	Multivariate logistic regression analysis based on personal and family history (affected 1st and 2nd degree relatives)	Proband: number of CRC, adenomas, endometrial cancer, other Lynch associated tumours, age at diagnosis of CRC, adenomas, endometrial cancer Family: CRC, endometrial cancer, other Lynch associated tumours, age at diagnosis for CRC and endometrial cancer in 1st and 2nd degree relatives AUC = 0.80, 95% CI 0.77 to 0.83
MMRpro[23]	Validation on 279 individuals from 226 clinic based families	Point mutations and large rearrangements at *MLH1, MSH2, MSH6*	Colon and endometrial	Mendelian and Bayesian analysis incorporating the penetrance and prevalence of mutations and predictive values of MSI and IHC. Accounts for family size and unaffected individuals	Input (proband and each 1st and 2nd degree relative): Relation to the proband, CRC and endometrial cancer status, age at diagnosis, current age or age at last follow-up if unaffected, MSI or IHC result, germline testing result of *MLH1, MSH2* or *MSH6* AUC = 0.83, 95% CI 0.78 to 0.88

AUC, area under the receiver operating characteristic curve; CI, confidence interval; CRC, colorectal cancer; IHC, immunohistochemistry; MSI, microsatellite instability.

CRC cohort for identification of *MLH1/MSH2/MSH6* mutation carriers.[22] Both models are based on logistic regression analyses that account for personal and family history and are available as web based tools. The MMRpro model[23] is a Bayesian tool that mimics an autosomal dominant pattern of inheritance based on parameters of prevalence of *MLH1/MSH2/MSH6* mutations and on the predictive values of tumour molecular characteristics. It considers the personal and family history of colorectal and endometrial cancer and data on MSI and is also available online free of charge.

Each of the models was developed in a different population and has not yet been directly compared with the others. The aim of our study was to compare the performance characteristics of the $PREMM_{1,2}$ and Barnetson models with clinical criteria and universal molecular screening in a population based cohort of CRC patients. Our major outcomes of interest were the ability of each strategy to identify individuals who should undergo tumour evaluation for MMR deficiency, as well as predict *MLH1/MSH2* mutation carriers.

PATIENTS AND METHODS

Patients

Between November 2000 and October 2001, all newly diagnosed CRC patients in 25 hospitals were enrolled in the EPICOLON study, whose goal was to establish the incidence of hereditary and familial CRC forms in Spain.[18 24] Exclusion criteria were familial adenomatous polyposis or patient or family refusal to participate in the study. The study was approved by the institutional ethics committee of each participating hospital, and written informed consent was obtained from all patients.

Demographic, clinical, and tumour related characteristics of probands, as well as a detailed family history were obtained using a pre-established questionnaire. Cancer histories in pedigrees were traced backward and laterally at least up to second degree relatives. Age at cancer diagnosis, type and location of the neoplasm and current status were recorded for each affected family member. Data on numbers and current ages of unaffected family members was not systematically collected.

Tumour microsatellite instability analysis and immunostaining

Microsatellite instability testing and immunostaining for MLH1/MSH2 proteins were performed in all patients regardless of age, personal or family history, and tumour characteristics. Microsatellite status was assessed using the five-marker panel proposed by the National Cancer Institute, as described elsewhere.[8] Tumours were classified as stable if none of the markers showed instability. Tumours with two or more unstable markers were classified as high level MSI (MSI-H) and tumours with one unstable marker were classified as low level MSI (MSI-L). Paraffin embedded sections were immunostained with antibodies against mismatch repair proteins (anti-MSH2, Oncogene Research Products, Boston, Massachusetts, USA; anti-MLH1, PharMingen, San Diego, California, USA).[18] Tumour cells were judged to be negative for protein expression only if they lacked staining in a sample in which normal colonocytes and stroma cells were stained.

MLH1/MSH2 germline mutation analysis

Patients found to have tumours with MMR deficiency (demonstrated by either MSI-H and/or lack of protein expression) underwent *MLH1/MSH2* germline genetic testing by both multiple ligation probe amplification (MLPA) analysis and sequencing, as previously described.[18] In addition, all individuals

Table 2 Overall distribution of patients fulfilling each strategy

Strategy	Number of patients fulfilling each strategy (n = 1222) (%)	Number of patients with mutations in *MLH1* or *MSH2* (n = 8) (%)	Number of patients with mutations missed by each strategy (n = 8) (%)
Amsterdam II	22 (2)	4 (50)	4 (50)
Any RBG	287 (23)	8 (100)	0
Universal MMR deficiency testing*	91 (7)	8 (100)	0
PREMM ⩾5%	396 (32)	8 (100)	0
PREMM ⩾10%	130 (11)	6 (75)	2 (25)
PREMM ⩾20%	32 (3)	5 (62)	3 (37)
PREMM ⩾40%	9 (<1)	2 (25)	6 (75)
Barnetson ⩾0.5%	349 (29)	7 (87)	1 (0.1)
Barnetson ⩾5%	77 (6)	6 (75)	2 (25)
Barnetson ⩾20%	29 (2)	4 (50)	4 (50)
Barnetson ⩾45%	14 (1)	3 (37)	5 (62)

RBG, revised Bethesda guidelines; MMR, mismatch repair.
*Considered as high microsatellite instability (MSI) or loss of protein expression by immunohistochemistry (IHC).

 J Med Genet 2008;45:557–563. doi:10.1136/jmg.2008.059311

with MMR proficient tumours and a PREMM$_{1,2}$ score ⩾20% (n = 18) were tested for germline mutations in *MLH1/MSH2* and for immunohistochemistry of MSH6 and PMS2 proteins.[25]

Statistical analysis

Values to each variable included in the equation of the PREMM$_{1,2}$ and the Barnetson models were assigned according to the personal and family history of each proband in the cohort. Performance characteristics including sensitivity, specificity, positive predictive value and negative predictive value with 95% confidence intervals (CI) were calculated with respect to the presence of unambiguous *MLH1/MSH2* germline mutations for the Amsterdam II criteria, RBG, molecular testing, PREMM$_{1,2}$ model and the Barnetson model. Because of the lack of available data on unaffected family members, we were not able to evaluate the performance of the MMRpro model in this cohort.

Sensitivity and specificity for the models depend on the cut-off used for the predicted risk of a mutation. Based on the original model development, we evaluated the following cut-off levels for the PREMM$_{1,2}$ model: ⩾5%, ⩾10%, ⩾20%, and ⩾40% and for the Barnetson model: ⩾0.5%, ⩾5%, ⩾20%, and ⩾45%.

Discrimination between patients with and without mutations was quantified by the area under the receiver operating characteristic curve (AUC) with 95% CI. Calibration was assessed by comparing the average predictions from each model to the observed prevalence of mutations. Calculations were performed using the SPSS software package 12.0 (SPSS Inc, Chicago, Illinois, USA).

RESULTS

Clinical characteristics of *MLH1/MSH2* mutation carriers

During the study period, 1222 patients with pathologically confirmed colorectal adenocarcinoma were diagnosed and included in the EPICOLON project. Demographic, clinical, and tumour related characteristics of patients included in the study have previously been described.[18] Overall, 91 patients (7%) had a tumour demonstrating MMR deficiency, defined as either MSI-H or loss of MLH1/MSH2 protein expression. Germline genetic testing identified eight (0.7%) deleterious mutations in *MSH2* (five cases) or *MLH1* (three cases) genes, and three variants of unknown significance (two in *MSH2* and one in *MLH1*). Neither *MLH1/MSH2* germline mutations nor loss of expression of MSH6/PMS2 proteins were found in the group of individuals with MMR proficient tumours and a PREMM$_{1,2}$ score ⩾20%.

Table 3 Distribution of patients and *MLH1/MSH2* mutation carriers according to risk groups by the PREMM$_{1,2}$ and the Barnetson models

	Number of patients (n = 1222) (%)	*MLH1/MSH2* mutation carriers (%)*
PREMM$_{1,2}$ score		
<5%	826 (68)	– (–)
5–9%	266 (22)	2 (0.8)
10–19%	98 (8)	1 (1)
20–39%	23 (2)	3 (13)
⩾40%	9 (0.7)	2 (22)
Barnetson score		
<0.5%	873 (71)	1 (0.1)
0.5–4%	272 (22)	1 (0.4)
5–19%	48 (3.7)	2 (4.2)
20–44%	15 (1.2)	1 (6.6)
⩾45%	14 (1)	3 (21)

*Percentages referred to each PREMM$_{1,2}$ or Barnetson category.

Clinical criteria, PREMM$_{1,2}$ and Barnetson scores

Distribution of individuals according to fulfilment of clinical criteria and to specified PREMM$_{1,2}$ and Barnetson model scores are reported in table 2. Fulfilment of any of the RBG, identification of MMR deficiency, or a PREMM$_{1,2}$ score ⩾5% identified all *MLH1/MSH2* mutation carriers. A Barnetson score ⩾0.5% missed one *MSH2* mutation carrier with a 0.3% predicted probability.

Overall, prevalence of mutation carriers according to risk groups by the PREMM$_{1,2}$ and the Barnetson models was lower than the predicted score of identifying a mutation in each pre-specified risk group (table 3). For the PREMM$_{1,2}$ model, distribution of mutation carriers was: <5%: 0; 5–9%: 2 (1%); 10–19%: 1 (1%); 20–39%: 3 (13%); ⩾40%: 2 (22%). For the Barnetson model, distribution of mutation carriers was: <0.5%: 1 (0.1%); 0.5–0.9%: 0; 1–4%: 1 (0.7%); 5–14%: 1 (2.5%); 15–19%: 1 (12.5%); 20–24%: 0; 25–34%: 0; 35–44%: 1 (14%); ⩾45%: 3 (21%).

Performance of predictive models

Performance characteristics of the PREMM$_{1,2}$ and the Barnetson models for the identification of *MLH/MSH2* gene carriers were analysed at the different cut-offs as presented in the original model development (table 4). The AUCs of the PREMM$_{1,2}$ and the Barnetson models were 0.93 (95% CI 0.86 to 0.99) and 0.92 (95% CI 0.83 to 1.01), respectively (fig 1). The average prediction of mutation carriers was 5% and 1% by the PREMM$_{1,2}$ and Barnetson models, respectively, compared to an observed prevalence of 0.7%.

The PREMM$_{1,2}$ model had 100% (95% CI 71% to 100%) sensitivity and 68% (95% CI 65% to 71%) specificity, when a 5% cut-off was used. Positive and negative predictive values were 2% (95% CI 1% to 4%) and 100% (95% CI 99% to 100%), respectively. No mutation carriers had a PREMM$_{1,2}$ score <5%. Sensitivity and specificity of increasing cut-offs of the PREMM$_{1,2}$ model were as follows: ⩾10%: 75% and 90%; ⩾20%: 62% and 98%; and ⩾40%: 25% and 99%, respectively.

The lowest cut-off analysed for the Barnetson model in the original study was 0.5%, which had 87% (95% CI 51% to 99%) sensitivity, 72% (95% CI 69% to 74%) specificity, 2% (95% CI 1% to 4%) predictive value, and 100% (95% CI 99% to 100%) negative predictive value in our cohort. Sensitivity and specificity of increasing cut-offs of the Barnetson model were: ⩾1%: 87% and 83%; ⩾5%: 75% and 94%; ⩾20%: 50% and 98%; ⩾45%: 37% and 99%.

Sensitivity of the RBG for identification of *MLH1/MSH2* mutation carriers was 100% (95% CI 71% to 100%), with 77% (95% CI 74% to 79%) specificity and a PPV of 3% (95% CI 1% to 5%). The Amsterdam II criteria missed four mutation carriers, with 50% (95% CI 22% to 78%) sensitivity, 98% (95% CI 97% to 99%) specificity and 18% PPV (95% CI 7% to 39%).

Performing MMR molecular screening in all tumours identified 91 patients with an MMR deficiency. This strategy had a sensitivity and specificity of 100% (95% CI 71% to 100%) and 94% (95% CI 92% to 95%), respectively, with a PPV of 10% (95% CI 5% to 18%).

Performance of predictive models combined with MMR deficiency

Combination of predictive scores with MMR deficiency increased specificity and PPV for each established cut-off (table 4). Consequently, a PREMM$_{1,2}$ score of ⩾5% with MMR deficiency had 97% (95% CI 96% to 98%) specificity and

Table 4 Performance characteristics for the identification of *MLH1/MSH2* mutation carriers according to different strategies

Strategy	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	With tumour MMR deficiency testing* Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)
Fulfilment of the revised Bethesda guidelines	100 (71 to 100)	77 (74 to 79)	3 (1 to 5)	100 (99 to 100)	98 (97 to 98)	22 (11 to 38)
Fulfilment of the Amsterdam I/II criteria	50 (22 to 78)	98 (97 to 99)	18 (7 to 39)	50 (22 to 78)	99 (99 to 100)	36 (15 to 65)
Universal MMR testing	–	–	–	100 (71 to 100)	94 (92 to 95)	10 (5 to 18)
PREMM⩾5%	100 (71 to 100)	68 (65 to 71)	2 (1 to 4)	100 (71 to 100)	97 (96 to 98)	21 (10 to 36)
PREMM⩾10%	75 (40 to 94)	90 (88 to 91)	5 (2 to 10)	75 (40 to 94)	99 (98 to 99)	29 (13 to 50)
PREMM⩾20%	62 (30 to 86)	98 (97 to 98)	16 (6 to 32)	62 (30 to 86)	99 (98 to 100)	36 (16 to 61)
PREMM⩾40%	25 (6 to 60)	99 (99 to 100)	22 (5 to 56)	25 (6 to 60)	99 (99 to 100)	29 (8 to 65)
Barnetson ⩾0.5%	87 (51 to 99)	72 (69 to 74)	2 (1 to 4)	87 (51 to 99)	98 (97 to 99)	21 (10 to 38)
Barnetson ⩾5%	75 (40 to 93)	94 (93 to 95)	8 (3 to 16)	75 (40 to 93)	99 (99 to 100)	35 (17 to 59)
Barnetson ⩾20%	50 (22 to 78)	98 (97 to 99)	14 (5 to 31)	50 (22 to 78)	100 (99 to 100)	44 (19 to 73)
Barnetson ⩾45%	37 (13 to 70)	99 (98 to 100)	21 (7 to 48)	37 (13 to 70)	100 (99 to 100)	37 (13 to 70)

CI, confidence interval; MMR, mismatch repair; NPV, negative predictive value; PPV, positive predictive value.
*Deficiency defined as high microsatellite instability (MSI) or loss of expression of MLH1 or MSH2.

21% (95% CI 10% to 36%) PPV. For the Barnetson cut off of ⩾0.5% combined with MMR deficiency, specificity increased to 98% (95% CI 97% to 99%) and PPV to 21% (95% CI 10% to 38%). Similarly, fulfilment of the RBG combined with MMR deficiency had 98% (95% CI 97% to 98%) specificity and 22% (95% CI 11% to 38%) PPV. Overall, PPV for each cut-off of the predictive models were substantially higher in combination with the presence of MMR deficiency in the tumour. When combined with MMR deficiency, PPV associated with higher model cutoffs were greater than the combination of fulfilment of the RBG and MMR deficiency (PPV of 29% for $PREMM_{1,2}$ ⩾10%, 35% for Barnetson ⩾5% and 22% for the RBG).

Characteristics of CRC patients with *MLH1/MSH2* mutations

Detailed personal and family histories of the eight mutation carriers are described in table 5. All mutation carriers but one were female and their median age at diagnosis was 64 years (28–81 years). Among mutation carriers, the predicted probability of identifying a mutation according to the $PREMM_{1,2}$ and the Barnetson models ranged from 5–89%, and from 0.3–84%, respectively. All mutation carriers fulfilled at least one of the RBG and showed MMR deficiency in their tumour.

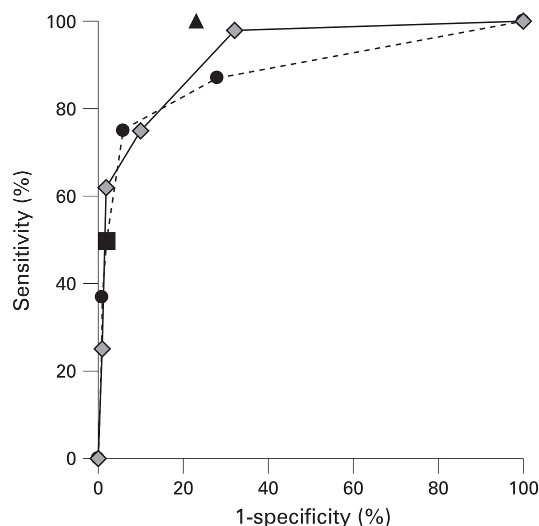


Figure 1 Sensitivity and 1-specificity receiver operating characteristic curves. The two lines illustrate the sensitivity and 1-specificity of the $PREMM_{1,2}$ model (—) and the Barnetson model (---). The square represents the sensitivity and 1-specificity value of the Amsterdam criteria and the triangle represents the sensitivity and 1-specificity value for the fulfilment of the revised Bethesda guidelines at single cutoff points, as these criteria are dichotomous.

DISCUSSION

We analysed the performance characteristics of the $PREMM_{1,2}$ and Barnetson models in a population based cohort of CRC patients for identifying *MLH1/MSH2* mutation carriers and compared them with clinical criteria and universal molecular screening for MMR deficiency with IHC and MSI. Because of the lack of available data on unaffected family members, we were not able to evaluate the performance of the MMRpro model in this cohort. The study shows that a ⩾5% cut off of the $PREMM_{1,2}$ model identified all *MLH1/MSH2* mutation carriers. The lowest threshold (0.5%) of the Barnetson model missed one mutation carrier. Tumour analysis for MMR deficiency improved the performance of the two predictive models at different cut-offs and yielded a similar likelihood of finding a mutation than the combination of fulfilling RBG and MMR deficiency at lower cutoffs and improved upon the RBG at higher thresholds.

Identification of appropriate individuals for genetic testing for Lynch syndrome is challenging due to insufficient specificity of current clinical criteria, the difficulty of implementing MMR deficiency screening for all CRC patients, the low prevalence of mutation carriers, and the high cost of genetic testing if performed in unselected cases. Selecting patients based on clinical criteria, presence of MMR deficiency, or a combination are all strategies that have been recommended by experts before ordering germline genetic testing.[17 18 26] According to our analysis, if one were to use the RBG as an initial selection strategy, 287 (23%) of CRC patients in this cohort would have been selected for further MMR deficiency testing. All *MLH1/MSH2* mutation carriers detected in this cohort would have been found. If a previous systematic quantitative approach were used such as ⩾5% score of the $PREMM_{1,2}$ model, 396 (32%) individuals would be selected to undergo molecular screening and no mutation carriers would have been missed. If the lowest cut-off originally analysed by the authors of the Barnetson model (⩾0.5%) was used, 349 individuals (29%) would have undergone further molecular testing, but one *MSH2* mutation carrier would have been missed with a 0.3% predicted probability. Use solely of the Amsterdam II criteria would have missed four mutation carriers (50% sensitivity), although there

Table 5 Predicted risks according to $PREMM_{1,2}$ and Barnetson models, fulfilment of clinical criteria, and molecular and clinical characteristics of patients with *MLH1/MSH2* mutations

ID	$PREMM_{1,2}$ score (%)	Barnetson score (%)	Revised Bethesda	AMS I/II	MSI status	Protein expression		Individual and familial clinical features*	Gene mutation
						MLH1	MSH2		
4037	5	0.3	Yes	No	MSI-H	Presence	Absence	Female (73), previous intestinal neoplasm (39), brother prostate cancer (71)	*MSH2*
15098	6	2	Yes	No	MSI-H	Absence	Presence	Female (49), 7 synchronous adenomas. No family history	*MLH1*
6040	13	5	Yes	No	MSI-H	Presence	Absence	Male (28). No family history	*MSH2*
4033	24	54	Yes	Yes	MSI-H	Absence	Presence	Female (47), father with CRC (39), sister with CRC (44)	*MLH1*
7047	26	38	Yes	No	MSI-H	Presence	Absence	Female (59), previous endometrial cancer (58), brother with CRC, sister with gynaecological cancer	*MSH2*
12084	36	18	Yes	Yes	MSI-H	Absence	Presence	Female (81), previous CRC/endometrial cancer (77), mother with CRC (49), brother with CRC (55), grandmother with endometrial cancer (78)	*MLH1*
12055	50	46	Yes	Yes	MSI-H	Presence	Absence	Female (69), previous CRC (53), mother with CRC (81), son with CRC (24), grandfather with gastric cancer (75), uncle with CNS tumour (55), and cousin with endometrial cancer (64)	*MSH2*
13071	89	84	Yes	Yes	MSI-H	Presence	Absence	76 year old female, synchronous adenoma in transversal colon, previous endometrial cancer (54), uncle with CRC (76), mother with endometrial cancer (47), brother with CRC (47), son with CRC (47), uncle with CRC (30)	*MSH2*

AMS, Amsterdam criteria; CNS, central nervous system; CRC, colorectal cancer; MSI, microsatellite instability; MSI-H: high degree of microsatellite instability.
*Age at diagnosis of the corresponding tumour in brackets.

would have been higher specificity and PPV (98% and 18%, respectively). Universal testing for MMR deficiency would require molecular screening of all 1222 cases to identify eight *MLH1/MSH2* mutation carriers.

It is important to emphasise that in the EPICOLON cohort, genetic testing for *MLH1/MSH2* was only performed in patients with tumour MMR deficiency, and theoretically, some mutation carriers whose tumours did not exhibit MMR deficiency may have been missed. In an effort to explore this possibility we tested all individuals with a $PREMM_{1,2}$ score ⩾20% (n = 18) for germline mutations in *MLH1/MSH2* and for immunohistochemistry of MSH6 and PMS2 proteins. Neither mutations nor loss of expression of the corresponding proteins were found. Ideally, the study that would most precisely answer the question of the true performance characteristics of each of the strategies for identification of mutation carriers would require comprehensive analysis of all four mismatch repair genes in a large population based cohort without tumour molecular prescreening for MMR deficiency; such a study has not thus far been undertaken.

Despite the above caveat, the EPICOLON cohort remains a useful population to externally validate the newly developed models. First, EPICOLON is entirely different from the moderate to high risk clinic population in which the $PREMM_{1,2}$ model was developed. Although it is also a population based cohort, because we included all ages of CRC, it is not as selected as the young onset series in which the Barnetson model was developed. Our analysis demonstrates that with sensitivities of 100%, the RBG or a $PREMM_{1,2}$ cutoff ⩾5% are both excellent options as tools for the initial risk assessment of CRC patients that may be used by a wide variety of clinicians in practice, including primary care physicians, oncologists, gastroenterologists or gynaecologists in deciding who should be referred for further genetic counselling and/or molecular evaluation. The Barnetson model also has excellent sensitivity but may miss a small proportion of mutation carriers if used as the sole method for this purpose. A formal cost effectiveness analysis would be extremely helpful to compare the costs associated with this surely more expensive and labour intensive strategy of universal IHC and/or MSI to identify mutation carriers than a more targeted approach. Of all CRC patients, 68–77% of individuals would not need any further molecular evaluation, and no mutations would have been missed, using the 5% $PREMM_{1,2}$ model threshold or the RBG, respectively.

The second question to address is which approach makes the best predictions of the likelihood of being a mutation carrier. The PPVs of the RBG, $PREMM_{1,2}$ ⩾5% and Barnetson ⩾0.5% were similar, at 3%, 2%, and 2%, respectively. While the two predictive models would initially select a slightly higher proportion of CRC cases than the RBG, they offer a quantitative risk assessment of finding a mutation and can be used to further refine the estimate of predicted probability, particularly in combination with MSI and IHC results. Performing molecular screening of MMR deficiency with MSI and IHC in all CRC was deemed to be sensitive and specific (100% and 94%), but had a PPV of only 10%. A combined finding of MMR deficiency and fulfilment of the RBG, $PREMM_{1,2}$ ⩾5% or Barnetson>0.5% all had similar PPV of 21–22%. At higher cutoffs, however, the models had improved PPVs compared to the RBG.

It is important to consider the limitations of the current validation of the predictive models. In this population based CRC cohort with a low overall mutation rate, both models had higher overall predictions compared to the actual observed rate, with $PREMM_{1,2}$ demonstrating poorer calibration. How the models function in other population based cohorts with higher mutation rates[27] and in higher risk clinical populations will be important to evaluate in future studies. It is also important to note that in their current forms, both the RBG and the Barnetson models are only applicable to patients with CRC, whereas $PREMM_{1,2}$ may also be used for risk assessment and mutation prediction in unaffected individuals and those with extracolonic tumours. Still, how the $PREMM_{1,2}$ model functions in these other patient groups needs to be further validated. The total number of mutation carriers in the EPICOLON cohort was small and germline testing for *MSH6* was not performed routinely. Although genetic testing was initially only performed in those patients with tumour MMR deficiency, it was subsequently performed in those cases with a $PREMM_{1,2}$ score

≥20% without MMR deficiency.[25] Neither mutation carriers nor cases with loss of expression of MSH6 or PMS2 were identified in this group. The low prevalence of mutations may also explain why the two models did not show a good calibration in this study cohort. Finally, comparison of the $PREMM_{1,2}$ and the Barnetson model with the MMRpro model[23] could not be performed due to lack of data on unaffected family members.

In conclusion, the $PREMM_{1,2}$ and the Barnetson models demonstrated similar performance to the RBG for identification of *MLH1/MSH2* mutation carriers in a population based cohort of CRC patients and allow quantitative risk assessment of the likelihood of finding an MMR gene mutation. Combination of the predictive models at specific cut-offs with tumour MMR deficiency selects those individuals more likely to carry an *MLH1/MSH2* mutation and offers a new clinical strategy to be considered for identification of individuals at risk of Lynch syndrome. Problems in calibration need to be addressed in different settings to warrant transportability of specific cut-offs into clinical decision making.

Funding: This work was supported by grants from the Fondo de Investigación Sanitaria (FIS 01/0104, 03/0070 and 05/0071), the Ministerio de Educación y Ciencia (SAF 04-07190 and SAF 07-64873) and the Asociación Española contra el Cáncer. Francesc Balaguer received a research grant from the Hospital Clínic and the Instituto de Salud Carlos III, and Sergi Castellví-Bel is supported by a contract from the Fondo de Investigación Sanitaria. CIBERehd is funded by the Instituto de Salud Carlos III. The work was also supported by the US National Cancer Institute grant CA 113433 (Dr Syngal).

Competing interests: None declared.

Ethics approval: The study was approved by the institutional ethics committee of each participating hospital.

Patient consent: Obtained.

REFERENCES

1. **Lynch HT,** de la Chapelle A. Genetic susceptibility to non-polyposis colorectal cancer. *J Med Genet* 1999;**36**:801–18.
2. **Giardiello FM,** Brensinger JD, Petersen GM. AGA technical review on hereditary colorectal cancer and genetic testing. *Gastroenterology* 2001;**121**:198–213.
3. **Aarnio M,** Sankila R, Pukkala E, Salovaara R, Aaltonen LA, de la Chapelle A, Peltomaki P, Mecklin J. P, Jarvinen H. J. Cancer risk in mutation carriers of DNA mismatch-repair genes. *Int J Cancer* 1999;**81**:214–8.
4. **Leach FS,** Nicolaides NC, Papadopoulos N, Liu B, Jen J, Parsons R, Peltomaki, P, Sistonen P, Aaltonen L. A, Nystrom-Lahti, M Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* 1993;**75**:1215–25.
5. **Papadopoulos N,** Nicolaides NC, Wei YF, Ruben SM, Carter KC, Rosen CA, Haseltine W. A, Fleischmann R. D, Fraser C. M, Adams, M. D. Mutation of a mutL homolog in hereditary colon cancer. *Science* 1994;**263**:1625–9.
6. **Miyaki M,** Konishi M, Tanaka K, Kikuchi-Yanoshita R, Muraoka M, Yasuno M, gari T, Koike M, Chiba M, Mori T. Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nat Genet* 1997;**17**:271–2.
7. **Chung DC,** Rustgi AK. The hereditary nonpolyposis colorectal cancer syndrome: genetics and clinical implications. *Ann Intern Med* 2003;**138**:560–70.
8. **Boland CR,** Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer S. J, Rodriguez-Bigas M. A, Fodde R, Ranzani G. 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**:5248–57.
9. **Jarvinen HJ,** Aarnio M, Mustonen H, Aktan-Collan K, Aaltonen LA, Peltomaki P, De La Chapelle A, Mecklin J. P Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer. *Gastroenterology* 2000;**118**:829–34.
10. **Schmeler KM,** Lynch HT, Chen LM, Munsell MF, Soliman PT, Clark MB, Daniels MS, White KG, Boyd-Rogers SG, Conrad PG, Yang KY, Rubin MM, Sun CC, Slomovitz BM, Gershenson DM, Lu KH. Prophylactic surgery to reduce the risk of gynecologic cancers in the Lynch syndrome. *N Engl J Med* 2006;**354**:261–9.
11. **Vasen HF,** Moslein G, Alonso A, Bernstein I, Bertario L, Blanco I, Burn, J. Capella G, Engel C, Frayling I, Friedl W, Hes F. J, Hodgson S, Mecklin J. P, Moller P, Nagengast F, Parc Y, Renkonen-Sinisalo L, Sampson J. R, Stormorken A, Wijnen J. Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer). *J Med Genet* 2007;**44**:353–62.
12. **Vasen HF,** Mecklin JP, Khan PM, Lynch HT. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). *Dis Colon Rectum* 1991;**34**:424–5.
13. **Vasen HF,** 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**:1453–6.
14. **Umar A,** Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff J, Fishel R, Lindor N.M, Burgart L.J, Hamelin R, Hamilton S.R, Hiatt R.A, Jass J, Lindblom A, Lynch H.T, Peltomaki P, Ramsey, S.D, Rodriguez-Bigas M.A, Vasen HF, Hawk ET, Barrett JC, Freedman AN, Srivastava S. Revised Bethesda guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst* 2004;**96**:261–8.
15. **Balmana J,** Stockwell DH, Steyerberg EW, Stoffel EM, Deffenbaugh AM, Reid JE, Ward B, Scholl T, Hendrickson B, Tazelaar J, Burbidge LA, Syngal S. Prediction of MLH1 and MSH2 mutations in Lynch syndrome. *JAMA* 2006;**296**:1469–78.
16. **Kievit W,** de Bruin JH, Adang EM, Ligtenberg MJ, Nagengast FM, van Krieken JH, Hoogerbrugge N. Current clinical selection strategies for identification of hereditary non-polyposis colorectal cancer families are inadequate: a meta-analysis. *Clin Genet* 2004;**65**:308–16.
17. **Hampel H,** Frankel WL, Martin E, Arnold M, Khanduja K, Kuebler P, Nakagawa H, Sotamaa K, Prior TW, Westman J, Panescu J, Fix D, Lockman J, Comeras I, de la Chapelle A. Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). *N Engl J Med* 2005;**352**:1851–60.
18. **Pinol V,** Castells A, Andreu M, Castellvi-Bel S, Alenda C, Llor X, Xicola RM, Rodriguez-Moranta F, Paya A, Jover R, Bessa X. Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. *JAMA* 2005;**293**:1986–94.
19. **Aaltonen LA,** Salovaara R, Kristo P, Canzian F, Hemminki A, Peltomaki P, Chadwick RB, Kaariainen H, Eskelinen M, Jarvinen H, Mecklin JP, de la Chapelle, A. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. *N Engl J Med* 1998;**338**:1481–7.
20. **Salovaara R,** Loukola A, Kristo P, Kaariainen H, Ahtola H, Eskelinen M, Harkonen N, Julkunen R, Kangas E, Ojala S, Tulikoura J, Valkamo E, Jarvinen H, Mecklin JP, Aaltonen LA, de la Chapelle A. Population-based molecular detection of hereditary nonpolyposis colorectal cancer. *J Clin Oncol* 2000;**18**:2193–200.
21. **Grover S,** Stoffel EM, Bussone L, Tschoegl E, Syngal S. Physician assessment of family cancer history and referral for genetic evaluation in colorectal cancer patients. *Clin Gastroenterol Hepatol* 2004;**2**:813–9.
22. **Barnetson RA,** Tenesa A, Farrington SM, Nicholl ID, Cetnarskyj R, Porteous ME, Campbell H, Dunlop MG. Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. *N Engl J Med* 2006;**354**:2751–63.
23. **Chen S,** Wang W, Lee S, Nafa K, Lee J, Romans K, Watson P, Gruber SB, Euhus D, Kinzler KW, Jass J, Gallinger S, Lindor NM, Casey G, Ellis N, Giardiello FM, Offit K, Parmigiani G. Prediction of germline mutations and cancer risk in the Lynch syndrome. *JAMA* 2006;**296**:1479–87.
24. **Pinol V,** Andreu M, Castells A, Paya A, Bessa X, Rodrigo J. Frequency of hereditary non-polyposis colorectal cancer and other colorectal cancer familial forms in Spain: a multicentre, prospective, nationwide study. *Eur J Gastroenterol Hepatol* 2004;**16**:39–45.
25. **Balaguer F,** Balmana J, Castellvi-Bel S, Steyerberg EW, Andreu M, Llor X, Jover R, Syngal S, Castells A. Validation and extension of the PREMM1,2 model in a population-based cohort of colorectal cancer patients. *Gastroenterology* 2008;**134**:39–46.
26. **Lynch HT,** Boland CR, Rodriguez-Bigas MA, Amos C, Lynch JF, Lynch PM. Who should be sent for genetic testing in hereditary colorectal cancer syndromes? *J Clin Oncol* 2007;**25**:3534–42.
27. **de la Chapelle A.** The incidence of Lynch syndrome. *Fam Cancer* 2005;**4**:233–7.

APPENDIX A

Investigators from the Gastrointestinal Oncology Group of the Spanish Gastroenterological Association who participated in the EPICOLON study

Hospital 12 de Octubre, Madrid: Juan Diego Morillas (local coordinator), Raquel Muñoz, Marisa Manzano, Francisco Colina, Jose Díaz, Carolina Ibarrola, Guadalupe López, Alberto Ibáñez; Hospital Clínic, Barcelona: Antoni Castells (local coordinator), Virgínia Piñol, Sergi Castellví-Bel, Francesc Balaguer, Victòria Gonzalo, Teresa Ocaña, Maria Pellisé, J. Ignasi Elizalde, Josep M. Piqué; Hospital Clínico Universitario, Zaragoza: Ángel Lanas (local coordinator), Javier Alcedo, Javier Ortego; Hospital Cristal-Piñor, Complexo Hospitalario de Ourense: Joaquin Cubiella (local coordinator), M♀ Soledad Díez, Mercedes Salgado, Eloy Sánchez, Mariano Vega; Hospital del Mar, Barcelona: Montserrat Andreu (local coordinator), Xavier Bessa, Agustín Panadés, Asumpta Munné, Felipe Bory, Miguel Nieto, Agustín Seoane; Hospital Donosti, San Sebastián: Luis Bujanda (local coordinator), Juan Ignacio Arenas, Isabel Montalvo, Julio Torrado, Ángel Cosme; Hospital General Universitario de Alicante: Artemio Payá (local coordinator), Rodrigo Jover, Juan Carlos Penalva, Cristina Alenda; Hospital General de Granollers: Joaquim Rigau (local coordinator), Ángel Serrano, Anna Giménez; Hospital General de Vic: Joan Saló (local coordinator), Eduard Batiste-Alentorn, Josefina Autonell, Ramon Barniol; Hospital General Universitario de Guadalajara: Ana María García (local coordinator), Fernando Carballo, Antonio Bienvenido, Eduardo Sanz, Fernando González, Jaime Sánchez; Hospital General Universitario de Valencia: Enrique Medina (local coordinator), Jaime Cuquerella, Pilar Canelles, Miguel Martorell, José Ángel García, Francisco Quiles, Elisa Orti; Hospital do Meixoeiro, Vigo: Juan Clofent (local coordinator), Jaime Seoane, Antoni Tardío, Eugenia Sanchez; Hospital San Eloy, Baracaldo: Luis Bujanda (local coordinator),

Carmen Muñoz, María del Mar Ramírez, Araceli Sánchez; Hospital Universitari Germans Trias i Pujol, Badalona: Xavier Llor (local coordinator), Rosa M. Xicola, Marta Piñol, Mercè Rosinach, Anna Roca, Elisenda Pons, José M. Hernández, Miquel A. Gassull; Hospital Universitari Mútua de Terrassa: Fernando Fernández-Bañares (local coordinator), Josep M. Viver, Antonio Salas, Jorge Espinós, Montserrat Forné, Maria Esteve; Hospital Universitari Arnau de Vilanova, Lleida: Josep M. Reñé (local coordinator), Carmen Piñol, Juan Buenestado, Joan Viñas; Hospital Universitario de Canarias: Enrique Quintero (local coordinator), David Nicolás, Adolfo Parra, Antonio Martín; Hospital Universitario La Fe, Valencia: Lidia Argüello (local coordinator), Vicente Pons, Virginia Pertejo, Teresa Sala; Hospital Universitario Reina Sofía, Córdoba: Antonio Naranjo (local coordinator), María del Valle García, Patricia López, Fernando López, Rosa Ortega, Javier Briceño, Javier Padillo; Fundació Hospital Son Llatzer, Palma de Mallorca: Àngels Vilella (local coordinator), Carlos Dolz, Hernan Andreu.

Raising the quality of academic journals

The Committee on Publication Ethics (COPE) is a charity that works to promote ethical conduct in scientific research and its publication in science journals. It was founded in 1997 by a group of medical journal editors concerned about research and publication misconduct. It now has over 3500 members, including many major medical and scientific journals. Its main activity is to advise the editors of academic journals on how to handle cases of suspected research and publication misconduct.

COPE is looking for new members to join its Council, which is responsible for the charity's policy and management. They will be expected to attend four meetings a year (in London) as well as take part in sub-committees and other COPE-related tasks. There are also occasional strategy meetings and an annual seminar to attend. Council members may also be asked to take responsibilities as trustees of the charity. This is a voluntary position. The initial term of office is three years, but members may be reappointed for a further term of three years. Applications from journal editors are particularly welcome.

Further details can be found on the website: www.publicationethics.org.uk

Those who are interested should send a short CV (**no more than 4 pages**) plus a covering letter stating what they can contribute to COPE Council. Please email to the COPE Administrator, Linda Gough (cope_administrator@publicationethics.org) or by post to Linda Gough, PO Box 39, Harleston, IP20 9WR, UK. The closing date for applications is **30 September 2008**.

Registered Charity No 1123023
Registered in England and Wales Company No 6389120

4. Discussion

The first study of this Doctoral Thesis (article #1) precisely quantities the relative importance of known clinical parameters in Lynch syndrome and develops a PRediction model for identification of *MLH1* and *MSH2* mutation carriers ($PREMM_{1,2}$ Model). The most significant clinical predictors of finding an *MLH1/MSH2* mutation according to proband's history were the presence of 2 or more colorectal cancers and according to family history were the number of first-degree relatives with colorectal or endometrial cancer. The $PREMM_{1,2}$ model was developed in a cohort of 1914 individuals at moderate risk of Lynch syndrome who provided a blood sample for genetic testing to Myriad Genetics Laboratories Inc. (Salt Lake City, UT). A development cohort included 898 unrelated probands and a validation cohort included 1016 additional unrelated probands. A multivariable logistic regression analysis considered personal and family history of the individual being tested for prediction of *MLH1* and *MSH2* mutations (both point mutations and large rearrangements). Variables included in the equation were: for the proband, presence, number and age at diagnosis of CRC, presence and age at diagnosis of endometrial cancer and colonic adenomas, and presence of other Lynch-associated neoplasms (urinary tract, gastric, small intestine, ovary, biliary tract, glioblastoma multiforme, and cutaneous sebaceous gland neoplasms); for the family, which considers first and second degree relatives, presence, number of first degree relatives and younger age at diagnosis of CRC and endometrial cancer, and presence of other Lynch-associated neoplasms. Secondly, robust coding of related predictors was performed; i.e., effects of age at diagnosis of CRC and endometrial were forced to be identical in probands, first and second degree relatives; similarly, CRC and endometrial diagnosis in 2nd degree relatives was weighted half that of 1st degree relatives[93]. The final and complex equation encompasses 12 variables

which were translated into a prediction rule presented in a user-friendly format for health care professionals at http://www.dfci.org/premm. (Figure 1)

Figure 1. $PREMM_{1,2}$ model as presented on the web

PREMM1,2 Model: Prediction Model for MLH1 and MSH2 Gene Mutations

The PREMM1,2 model is a clinical prediction rule designed to be used by healthcare professionals to estimate the probability that an individual carries a mutation in *MLH1* or *MSH2*. Mutations in these genes are found in most patients with the Lynch syndrome. Learn more

Proband Information
("Proband" refers to the individual being evaluated. Ideally, this individual should have a cancer or adenoma diagnosis.)

How many separate colorectal cancers has the proband had?

- None
- One
- Two or more

If one, what was the age at diagnosis?
(if unknown, estimate)

If two or more, what was the youngest age at diagnosis?
(if unknown, estimate)

Has the proband had colonic adenoma(s)?

- Yes
- No

What was the youngest age at diagnosis?
(if unknown, estimate)

Has the proband had endometrial cancer?

- Yes
- No

What was the youngest age at diagnosis?
(if unknown, estimate)

Has the proband had another HNPCC-associated cancer?

- Yes
- No

(Other HNPCC-associated cancers include ovary, stomach, small intestine, urinary tract/kidney, bile ducts, glioblastoma multiforme, sebaceousgland tumors, and pancreas.)

Relatives Information - First Degree
(Only from affected side of family)

How many first-degree relatives
have had colorectal cancer?

- None
- One
- Two or more

If one, what was his/her age at diagnosis?

(if unknown, estimate)
If two or more, what was the youngest age at diagnosis?

(if unknown, estimate)
How many first-degree relatives
have had endometrial cancer?

- None
- One
- Two or more

If one, what was her age at diagnosis?

(if unknown, estimate)
If two or more, what was the youngest age at diagnosis?

(if unknown, estimate)
Have any first-degree relatives
had another HNPCC-associated cancer?

- Yes
- No

Relatives Information - Second Degree
(Only from affected side of family)

How many second-degree relatives
have had colorectal cancer?

- None
- One
- Two or more

If one, what was his/her age at diagnosis?

(if unknown, estimate)
If two or more, what was the youngest age at diagnosis?

(if unknown, estimate)
How many second-degree relatives
have had endometrial cancer?

- None
- One
- Two or more

Have any second-degree relatives
had another HNPCC-associated cancer?

Yes

No

calculate

Predicted Probability of Mutation

The PREMM$_{1,2}$ model is a quantitative tool to estimate the likelihood of carrying a mutation in the *MLH1* or *MSH2* genes based on simple features from the personal and family history of an individual. The model was turned into a user-friendly format to allow health care professionals quantify the likelihood of an individual carrying a mutation in any of these genes. The model provides individual's risk for each member of a family and therefore may help to determine which family member might be most appropriate for testing. How the risk prediction is translated into clinical practice depends on many other factors, such as ability to pursue tumor molecular screening, availability of comprehensive genetic services, or the urge to obtain a genetic result for clinical decision making. Based on the risk and some of the above factors one may decide to refer the individual for specialized risk assessment, pursue molecular screening, or decide for direct germline genetic testing. In addition, the risk quantification eases the communication process with the patient, as it may facilitate the understanding of not pursuing a molecular work-up if the estimated risk of identifying a mutation is very low.

The PREMM$_{1,2}$ model showed an excellent calibration and an area under the receiving operating curve (AUC) of 0.80 (95% CI 0.77-0.83). Nevertheless, the PREMM$_{1,2}$ model has some limitations in its current format. It does not include

prediction of *MSH6*, and MSI and IHC data were not incorporated as molecular predictors of a germline mutation in the current version. In addition, it does not consider family size or unaffected family members. Some of these limitations are currently being addressed in the expansion of the $PREMM_{1,2,6}$ model.

Initially, the MMRpredict, MMRpro and $PREMM_{1,2}$ models were each compared in their original development or validation cohorts against current clinical criteria used for identification of individuals with suspicion of Lynch syndrome to assess their performance. In their own series, performance of the MMRpredict model at a cutoff of 0.005 was similar to the Bethesda criteria. The model performed better than the Amsterdam criteria either alone or in combination with tumor MMR testing[90]. In the original cohort, comparison of the sensitivity and specificity of the $PREMM_{1,2}$ model with the revised Bethesda guidelines depended on the cutoff used for the predicted risk of a mutation. A 5% cut off selected 20% more individuals for genetic testing than the revised Bethesda guidelines (higher sensitivity), with 10% lower specificity, but had a lower miss rate of mutation carriers, 6% versus 26%. Finally, validation of the MMRpro model in 226 clinic-based families resulted in higher accuracy than current clinical criteria to identify mutation carriers based on its higher discriminatory ability[91].

Therefore, after all these recent models were published there was the need to compare all of them in the same study population and in different settings from the original in order to assess their performance and to warrant transportability, respectively. This is the goal we aimed to achieve with the second article of this Thesis.

The EPICOLON cohort is a Spanish population-based cohort of 1222 consecutive CRC individuals whose tumors were tested for MMR deficiency[94]. In this cohort, those individuals with MSI or loss of expression of MLH1 and/or MSH2 by IHC proceeded to germline analysis. Overall, 91 patients (7%) had a tumor demonstrating

MMR deficiency, defined as either MSI-H or loss of MLH1/MSH2 protein expression. Germline genetic testing identified eight (0.7%) deleterious mutations in *MSH2* (5 cases) and *MLH1* (3 cases).[28]

The second study of the Thesis (article #2) was focused on validating and comparing the $PREMM_{1,2}$ model with other clinical strategies in the EPICOLON cohort. The $PREMM_{1,2}$ and MMRpredict models for identification of *MLH1/MSH2* mutation carriers were compared with the revised Bethesda guidelines, Amsterdam II criteria, and tumor analysis for MMR deficiency in the EPICOLON cohort. This work represented the first analysis of comparison in the same study population of two recent developed models with current clinical strategies for identification of *MLH1/MSH2* mutation carriers. The major outcome of interest was to evaluate the ability of each strategy to identify individuals who should undergo tumor evaluation for MMR deficiency, as well as predict *MLH1/MSH2* mutation carriers. Interestingly, both models had a similar discriminatory ability with an AUC of 0.93 (95% CI 0.86 to 0.99) for $PREMM_{1,2}$ and 0.92 (95% CI 0.83 to 1.01) for MMRpredict. The analyses demonstrated that with sensitivities of 100%, the revised Bethesda guidelines or a $PREMM_{1,2}$ score ≥ 5% are both excellent options as tools for the initial risk assessment of CRC patients that may be used by a wide variety of clinicians in practice in deciding who should be referred for further genetic counseling and/or molecular evaluation. The MMRpredict model also has an excellent sensitivity but may miss a small proportion of mutation carriers if used as the sole method for this purpose (87% sensitivity, as it missed one mutation carrier with a predicted probability below 0.5%). The Amsterdam II criteria missed four mutation carriers (50% sensitivity), and universal testing for MMR deficiency required molecular screening of all 1222 cases to identify 8 *MLH1/MSH2* mutation carriers. This strategy had a sensitivity and specificity of 100% and 94%, respectively, with a 10%

PPV. Nevertheless, this strategy is likely to be more expensive and labor intensive than a more targeted approach. In fact, of all CRC patients from the EPICOLON cohort, 68-77% of individuals would not need any further molecular evaluation, and no mutations would have been missed, using the 5% $PREMM_{1,2}$ model threshold or the revised Bethesda guidelines, respectively.

The second question to address is which approach makes the best predictions of the likelihood of being a mutation carrier. The positive predictive values (PPV) of the revised Bethesda guidelines, $PREMM_{1,2}$ ≥5% and MMRpredict ≥0.5% were similar, at 3%, 2%, and 2% respectively. While the two predictive models would initially select a slightly higher proportion of CRC cases than the revised Bethesda guidelines, they offer a quantitative risk assessment of finding a mutation and can be used to further refine the estimate of predicted probability, particularly in combination with MSI and IHC results. Performing molecular screening of MMR deficiency with MSI and IHC in all CRC was deemed to be sensitive and specific (100% and 94%), but had a PPV of only 10%. A combined finding of MMR deficiency and fulfillment of the revised Bethesda guidelines, $PREMM_{1,2}$ ≥5% or MMRpredict>0.5% all had similar PPV of 21-22%. At higher cutoffs, however, the models had improved PPVs compared to the revised Bethesda guidelines.

Taking into account these results, we proposed a strategy for clinical application of the $PREMM_{1,2}$ model in patients with CRC. If the individual were to obtain a low prediction score (such as less than 5%), no further molecular analysis would be indicated. If the individual had a score between 5 and 19%, tumor MMR testing would be indicated, and if the individual reached a score equal or higher than 20%, direct germline genetic testing could be considered if tumor sample were not available[95].

The current validation of the predictive models, and specially the $PREMM_{1,2}$ model, in the EPICOLON cohort must be considered in light of some limitations. First, it only included CRC patients. While the revised Bethesda guidelines and the MMRpredict model were designed to identify CRC patients at risk of Lynch syndrome, the $PREMM_{1,2}$ model may also be used for risk assessment and mutation prediction in unaffected individuals and those with extracolonic tumors. Nevertheless, how the $PREMM_{1,2}$ model functions in these other patient groups needs to be further validated. The total number of mutation carriers in the EPICOLON cohort was small and germline testing for *MSH6* was not performed routinely. Finally, comparison of the $PREMM_{1,2}$ and the MMRpredict model with the MMRpro model could not be performed due to lack of data on unaffected family members. Consequently, external validation is currently ongoing in an international collaboration that gathers individuals with complete family history who have been tested for Lynch syndrome in high-risk clinics and population-based.

In conclusion, results from the second article suggests that the $PREMM_{1,2}$ model might be a good discriminatory and easy-to-use clinical tool to help health care professionals select those individuals with CRC who may benefit from further molecular testing to find out if they carry a *MLH1* or *MSH2* mutation. Prediction estimates in combination with MMR deficiency testing may further help delineate which individuals are more likely to have a mutation and offers a new clinical strategy to be considered for identification of individuals at risk of Lynch syndrome.

5. Conclusions

Results obtained from the studies presented in this Thesis allow us to conclude:

- Personal and family history characteristics can accurately predict the outcome of genetic testing in a large population at risk of Lynch syndrome.
- Strong predictors of an *MLH1/MSH2* mutation included proband characteristics such as presence of colorectal cancer, especially if more than 2 separate diagnoses, or endometrial cancer; and family history such as the number of first-degree relatives with colorectal or endometrial cancer. Age at diagnosis was important for colorectal cancer.
- The $PREMM_{1,2}$ model provides clinicians with an easy-to-use tool to estimate the likelihood of finding a mutation in the *MLH1/MSH2* genes.
- The $PREMM_{1,2}$ model is useful to identify *MLH1/MSH2* mutation carriers among unselected colorectal cancer patients.
- The $PREMM_{1,2}$ and the MMRpredict models perform similar to the revised Bethesda guidelines to select CRC patients for identification of *MLH1/MSH2* mutation carriers.
- Among CRC patients, high predictions of the $PREMM_{1,2}$ and the MMRpredict models in combination with MMR deficiency perform better than fulfillment of the revised Bethesda guidelines with MMR deficiency.
- The $PREMM_{1,2}$ and the MMRpredict models offer a quantitative assessment of the genetic risk that might be useful to decide on patient's referral, subsequent tumor evaluation and germline genetic testing.

6. Next steps

The $PREMM_{1,2}$ model provides an overall quantitative estimation of the probability of finding an *MLH1/MSH2* mutation in the individual being evaluated based on clinical information. A step forward is to expand the $PREMM_{1,2}$ model to incorporate prediction of finding an *MSH6* mutation, as well as obtain specific estimates for each gene separately. Moving on, incorporation of molecular data such as MMR deficiency (MSI or loss of protein expression) into the model may help refine the accuracy of the estimates. In addition, since the issue for researchers and clinicians is whether we are ready to incorporate the $PREMM_{1,2}$ model and the others into clinical practice and to define their role in the identification of Lynch syndrome individuals, further validation studies in different settings and study populations are needed to evaluate each model cutoff performance with respect to clinical decision making. Further validation analysis in different population and clinic-based cohorts from CRC or endometrial cancer patients will help to refine their clinical application and are warranted for a more generalized transportability. Finally, a cost-effectiveness analysis comparing fulfillment of clinical criteria, universal molecular testing, a determined cut off by each of the predictive models, or a combination of the above strategies may offer a new approach for public health decisions.

7. Appendix

7.1. Appendix 1. "Validation and extension of the $PREMM_{1,2}$ Model in a population-based cohort of colorectal cancer patients".

Francesc Balaguer, Judith Balmaña, Sergi Castellví-Bel, Ewout W. Steyerberg, Montserrat Andreu, Xavier Llor, Rodrigo Jover, Sapna Syngal, Antoni Castells for the Gastrointestinal Oncology Group of the Spanish Gastroenterological Association. Gastroenterology 2008; 134:39-46

GASTROENTEROLOGY 2008;134:39–46

CLINICAL–ALIMENTARY TRACT

Validation and Extension of the $PREMM_{1,2}$ Model in a Population-Based Cohort of Colorectal Cancer Patients

FRANCESC BALAGUER,* JUDITH BALMAÑA,‡ SERGI CASTELLVÍ–BEL,* EWOUT W. STEYERBERG,§ MONTSERRAT ANDREU,‖ XAVIER LLOR,¶ RODRIGO JOVER,# SAPNA SYNGAL,‡‡ and ANTONI CASTELLS* for the Gastrointestinal Oncology Group of the Spanish Gastroenterological Association‡‡

*Department of Gastroenterology, Institut de Malalties Digestives i Metabòliques, Hospital Clínic, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, Institut d'Investigacions Biomèdiques August Pi i Sunyer, University of Barcelona, Barcelona, Catalonia, Spain; ‡Department of Medical Oncology, Hospital Vall d'Hebron, Barcelona, Catalonia, Spain; §Department of Public Health, Erasmus Medical Center–University MC Rotterdam, the Netherlands; ‖Department of Gastroenterology, Hospital del Mar, Barcelona, Catalonia, Spain; ¶Department of Gastroenterology, Hospital Universitari Germans Trias i Pujol, Badalona, Catalonia, Spain; #Department of Gastroenterology, Hospital General Universitario de Alicante, Alicante, Spain; and the ‡‡Division of Gastroenterology, Brigham and Women's Hospital and Population Sciences Division, Dana-Farber Cancer Institute, Boston, Massachusetts

Background & Aims: Early recognition of patients at risk for Lynch syndrome is critical but often difficult. Recently, a predictive algorithm—the $PREMM_{1,2}$ model—has been developed to quantify the risk of carrying a germline mutation in the mismatch repair (MMR) genes *MLH1* and *MSH2*. However, the model's performance in an unselected, population-based colorectal cancer population as well as its performance in combination with tumor MMR testing are unknown. Methods: We included all colorectal cancer cases from the EPICOLON study, a prospective, multicenter, population-based cohort (n = 1222). All patients underwent tumor microsatellite instability analysis and immunostaining for MLH1 and MSH2, and those with MMR deficiency (n = 91) underwent tumor *BRAF* V600E mutation analysis and *MLH1/MSH2* germline testing. Results: The $PREMM_{1,2}$ model with a ≥5% cut-off had a sensitivity, specificity, and positive predictive value (PPV) of 100%, 68%, and 2%, respectively. The use of a higher $PREMM_{1,2}$ cut-off provided a higher specificity and PPV, at expense of a lower sensitivity. The combination of a ≥5% cut-off with tumor MMR testing maintained 100% sensitivity with an increased specificity (97%) and PPV (21%). The PPV of a $PREMM_{1,2}$ score ≥20% alone (16%) approached the PPV obtained with $PREMM_{1,2}$ score ≥5% combined with tumor MMR testing. In addition, a $PREMM_{1,2}$ score of <5% was associated with a high likelihood of a *BRAF* V600E mutation. Conclusions: The $PREMM_{1,2}$ model is useful to identify *MLH1/MSH2* mutation carriers among unselected colorectal cancer patients. Quantitative assessment of the genetic risk might be useful to decide on subsequent tumor MMR and germline testing.

Lynch syndrome, also called hereditary nonpolyposis colorectal cancer, is the most common form of hereditary colorectal cancer (CRC), accounting for 1% to 5% of all colorectal malignancies.[1–3] It is characterized by early onset of CRC and other adenocarcinomas, predominantly endometrial cancer. The syndrome is inherited in an autosomal dominant pattern with variable penetrance and occurs as a consequence of germline mutations in the mismatch repair (MMR) system,[4] mainly in *MLH1* and *MSH2* (>90% of cases)[1] but also in *MSH6*[5] and *PMS2*.[6] The abnormal function of these genes leads to the accumulation of errors during DNA replication, particularly in repetitive sequences (microsatellites). As a result, tumors in patients with Lynch syndrome characteristically demonstrate microsatellite instability (MSI)[7] as well as loss of protein expression corresponding to the mutated gene.[8]

The heterogeneity of Lynch syndrome complicates early recognition, which is critical and often not straightforward. The diagnostic criteria continue to evolve as understanding and characterization of this disorder improve. Indeed, identification of Lynch syndrome can be done by tumor MMR screening using MSI testing and/or immunostaining, in combination or not with clinical criteria. At present, the most widely accepted strategy relies on tumor molecular analysis in patients fulfilling the revised Bethesda guidelines.[7] Nevertheless, as in hereditary breast-ovarian cancer syndrome in the past,[9,10] Lynch syndrome identification is moving toward more refined algorithms and multivariable models that com-

Abbreviations used in this paper: CRC, colorectal cancer; MLPA, multiple ligation probe amplification; MMR, mismatch repair; MSI, microsatellite instability; PPV, positive predictive value.

© 2008 by the AGA Institute
0016-5085/08/$34.00
doi:10.1053/j.gastro.2007.10.042

bine personal and familial data to obtain a quantitative estimation of the risk.[11–14]

The $PREMM_{1,2}$ model[11] is a recently developed Web-based logistic regression model that predicts the likelihood of germline mutations in the *MLH1* and *MSH2* genes on the basis of personal and family history of individuals. It was developed in a large and diverse cohort of probands undergoing genetic testing on the basis of their clinical history. Whereas the model accurately discriminates gene mutation carriers in this subset of individuals at moderate to high risk for Lynch syndrome,[11] its usefulness in an unselected CRC population is unknown. Furthermore, efficacy of the $PREMM_{1,2}$ model in combination with tumor MMR testing has not yet been assessed.

Using data from the EPICOLON study[15,16]—a prospective, multicenter, population-based cohort collected to establish the incidence and characteristics of hereditary and familial CRC forms in Spain—we assessed the efficacy of the $PREMM_{1,2}$ model, in combination or not with tumor MMR testing, for the identification of *MLH1* and *MHS2* gene mutation carriers among unselected CRC patients.

Materials and Methods

Patients

Between November 2000 and October 2001, all newly diagnosed CRC patients in 25 hospitals were included in the EPICOLON study.[15,16] Exclusion criteria were familial adenomatous polyposis, personal history of inflammatory bowel disease, and patient or family refusal to participate in the study. The study was approved by the institutional ethics committee of each participating hospital, and written informed consent was obtained from all patients.

Demographic, clinical, and tumor-related characteristics of probands, as well as a detailed family history were obtained using a pre-established questionnaire. Pedigrees were traced backward and laterally as far as possible, or at least up to second-degree relatives, in terms of cancer history. Age at cancer diagnosis, type, location, and tumor stage of the neoplasm and current status were recorded for each affected family member.[15,16]

Tumor Microsatellite Instability Analysis and Immunostaining

Tissue samples from tumor and normal colonic mucosa were obtained from each patient, immediately frozen in liquid nitrogen, and stored at −70°C until use. In cases where no frozen tissue was available, formalin-fixed, paraffin-embedded samples were used. Genomic DNA was isolated using the QiaAmp Tissue Kit (Qiagen, Courtaboeuf, France).

Microsatellite instability testing and immunostaining for MLH1 and MSH2 were performed in all patients regardless of age, personal or family history, and tumor characteristics. In addition, in those patients with a $PREMM_{1,2}$ score ≥20%, immunostaining for MSH6 and PMS2 was also performed. Paraffin-embedded sections were immunostained with antibodies against mismatch repair proteins (anti-MSH2, Oncogene Research Products, Boston, MA; anti-MLH1, PharMingen, San Diego, CA; anti-MSH6, BD Transduction Laboratories; anti-PMS2, PharMingen), as described elsewhere.[15] Tumor cells were judged to be negative for protein expression only if they lacked staining in a sample in which normal colonocytes and stroma cells were stained. If no immunostaining of normal tissue could be demonstrated, the results were considered ambiguous.

Microsatellite status was assessed using the 5-marker panel proposed by the National Cancer Institute, as described elsewhere.[15,17,18] Tumors were classified as stable if none of the markers showed instability. Tumors with 2 or more unstable markers were classified as high level MSI (MSI-H) and tumors with 1 unstable marker were classified as low-level MSI (MSI-L).

Germline MLH1/MSH2 Mutation Analysis

Patients found to have tumors with MMR deficiency (demonstrated by either MSI-H and/or lack of protein expression) underwent *MSH2/MLH1* germline genetic testing. Moreover, all patients with a $PREMM_{1,2}$ score ≥20% with MMR-proficient tumors also underwent genetic testing.

Germline mutational analysis was performed by both multiple ligation probe amplification (MLPA) analysis and sequencing, as described elsewhere.[15]

Tumor BRAF V600E Mutation Analysis

Tumor *BRAF* V600E mutation analysis was performed in all patients with MSI (high and low) and/or lack of MLH1/MSH2 protein expression by direct sequencing in tumor DNA, as described elsewhere.[19]

Application of the $PREMM_{1,2}$ Model

The $PREMM_{1,2}$ model is a clinical model created to predict the likelihood of finding a *MLH1* or *MSH2* mutation in at-risk individuals.[11] The original study analyzed *MLH1/MSH2* mutation prevalence in a large cohort of patients undergoing genetic testing at Myriad Genetic Laboratories Inc (Salt Lake City, UT). A multivariable model using logistic regression and including variables related to the proband and relatives was developed. The prediction rule is available as a Web-based tool at the Dana-Farber Cancer Institute web site (http://www.dfci.org/premm). We calculated the $PREMM_{1,2}$ score for each patient included in the study using the SPSS V11.0 software package (SPSS Inc, Chicago, IL).

Statistical Analysis

Sensitivity, specificity, and positive predictive value (PPV) of the $PREMM_{1,2}$ model, either alone or in

Table 1. Clinical and Molecular Characteristics of Patients Included in the Study

Characteristics	Values
Age (*y*)[a]	70 ± 11
Men: no. (%)	731 (59.8)
Site of tumor: no. (%)	
Proximal to splenic flexure	357 (29.2)
Distal to splenic flexure	865 (70.8)
Tumor TNM stage: no. (%)	
I	161 (13.2)
II	510 (41.7)
III	337 (27.6)
IV	214 (17.5)
Degree of differentiation: no. (%)	
Well	292 (23.9)
Moderate	835 (68.3)
Poor	95 (7.8)
Mucinous carcinoma type: no. (%)	142 (11.6)
Synchronous colorectal cancer: no. (%)	71 (5.8)
Microsatellite instability, high: no. (%)	83 (6.8)
Microsatellite instability, low: no. (%)	28 (2.3)
Loss of MLH1/MSH2 protein expression: no. (%)	81 (6.6)
Tumor MMR deficiency[b]: no. (%)	91 (7.4)
Germline *MLH1* mutation: no. (%)	3 (0.25)
Germline *MSH2* mutation: no. (%)	5 (0.4)

MMR, mismatch repair.
[a]Expressed as mean ± standard deviation.
[b]Tumor MMR deficiency demonstrated by either high microsatellite instability and/or loss of MLH1/MSH2 protein expression.

combination with tumor MMR testing, were calculated with respect to the presence of *MLH1/MSH2* germline mutations. These performance characteristics depend on the cut-off used for the predicted risk of mutation, and we therefore arbitrarily evaluated the following cut-off levels: <5%, ≥5%, ≥10%, ≥20%, and ≥40%. Ninety-five percent binomial confidence intervals were calculated on the basis of the Adjusted Wald method.[20]

Continuous variables were expressed as mean ± standard deviation and compared by the Student *t* test. Categorical variables were compared by the χ^2 test, applying the Yates correction when needed.

All *P* values were two sided. A *P* value of less than .05 was considered to indicate a statistically significant difference. All calculations were performed using the 11.0 SPSS software package (SPSS Inc).

Results

Characteristics of the Patients

During the study period, 1222 patients with pathologically confirmed colorectal adenocarcinoma were diagnosed and included in the EPICOLON project. Demographic, clinical, and tumor-related characteristics of patients included in the study are summarized in Table 1.

One hundred eleven (9.1%) patients showed tumor MSI, 83 of them (6.8%) were MSI-H and 28 (2.3%) were MSI-L. Likewise, 81 (6.6%) patients had a tumor with loss of protein expression in either MLH1 (60 cases) or MSH2 (21 cases). No patients with tumors that were MSI-L had lack of MMR protein expression. However, expression of both proteins was retained in 10 tumors with MSI-H, whereas loss of MLH1 or MSH2 expression was found in 8 patients whose tumor did not show MSI. Overall, 91 (7.4%) patients were found to have a tumor demonstrating MMR deficiency (defined as MSI-H and/or loss of MLH1 or MSH2 expression).

BRAF V600E mutation was detected in 20 of 83 (24.1%) MSI-H tumors and in 18 of 60 (30%) tumors exhibiting loss of MLH1 expression. In addition, only 2 of 28 (7.1%) MSI-L tumors showed the *BRAF* V600E mutation, whereas it was not observed in any tumor with loss of MSH2 expression.

Germline genetic testing identified 8 (0.7%) unambiguous mutations in either *MSH2* (5 cases) or *MLH1* (3 cases) genes.

Efficacy of the PREMM$_{1,2}$ Model for the Identification of MLH1/MSH2 Gene Carriers

The distribution according to the PREMM$_{1,2}$ predicted likelihood of carrying a *MLH1/MSH2* germline mutation in the cohort was: <5%, 826 (68%); 5–9%, 266 (22%); 10–19%, 98 (8%); 20–29%, 23 (2%); and >40%, 9 (0.7%).

We first evaluated the PREMM$_{1,2}$ model for its ability to identify *MLH1/MSH2* mutation carriers within the large cohort of CRC patients. Performance characteristics of the PREMM$_{1,2}$ model for the identification of *MLH1/MSH2* gene carriers depended on the cut-off used for the predicted risk of mutation (Table 2). Using a cut-off of

Table 2. Performance Characteristics of the PREMM$_{1,2}$ Model for the Identification of *MLH1/MSH2* Gene Mutation Carriers

	Without tumor MMR test results[a]				With MSI-H or abnormal IHC tumor results[a]			
Strategy	No. (%)	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	No. (%)	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)
PREMM$_{1,2}$ ≥5%	396 (32.0)	100 (70.6-100)	68 (65.4-70.6)	2 (1-4)	39 (3.2)	100 (70.6-100)	97.4 (96.4-98.2)	20.5 (10.5-35.8)
PREMM$_{1,2}$ ≥10%	130 (10.6)	75 (40-93.7)	89.8 (88-91.4)	4.6 (1.9-9.9)	21 (1.7)	75 (40-93.7)	98.8 (97.9-99.3)	28.6 (13.5-50.2)
PREMM$_{1,2}$ ≥20%	32 (2.6)	62.5 (30.4-86.5)	97.7 (96.8-98.5)	15.6 (6.4-32.2)	14 (1.1)	62.5 (30.4-86.5)	99.3 (98.6-99.6)	35.7 (16.2-61.4)
PREMM$_{1,2}$ ≥40%	9 (0.7)	25 (6.3-59.9)	99.4 (98.8-99.7)	22.2 (5.3-55.7)	7 (0.6)	25 (6.3-59.9)	99.6 (99-99.8)	28.6 (7.6-64.8)

MMR, mismatch repair; MSI-H, high microsatellite instability; IHC, immunohistochemistry; 95% CI, 95% confidence interval.
[a]Tumor MMR testing by either microsatellite instability or *MLH1/MSH2* immunostaining.

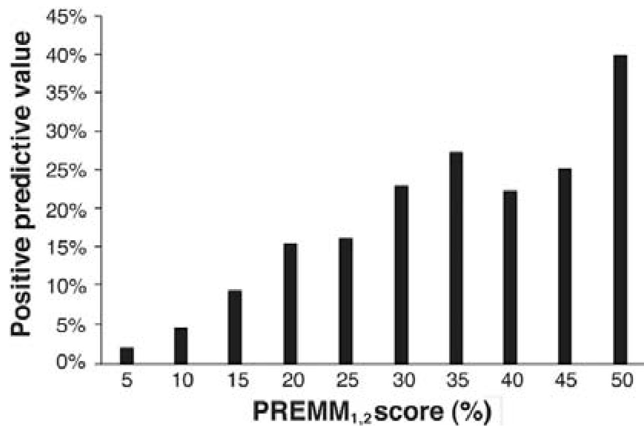


Figure 1. Positive predictive value for detecting germline *MLH1/MSH2* gene mutations according to the $PREMM_{1,2}$ score.

≥5%, the model had a sensitivity of 100%; therefore, no mutation carriers would be missed if molecular evaluation was restricted to individuals with a $PREMM_{1,2}$ score of ≥5%. Using higher cut-offs of 10%, 20%, and 40% led to a progressive loss of sensitivity (75%, 62.5%, and 25%, respectively). As expected, specificity increased with higher cut-offs and ranged from 68% with a 5% cut-off to 99.4% with a 40% cut-off. Positive predictive values of different cut-offs for the $PREMM_{1,2}$ model are depicted in Figure 1.

Use of the $PREMM_{1,2}$ Model in Combination With Tumor MMR Testing

The addition of tumor MMR testing, either by MSI analysis or immunostaining, to the $PREMM_{1,2}$ model enhanced its performance by improving both specificity and PPV (Table 2). A $PREMM_{1,2}$ score of ≥5% in combination with abnormal MMR testing was associated with a sensitivity of 100%, specificity of 97.4%, and PPV of 20.5%. The maximum PPV (36%) was achieved using a $PREMM_{1,2}$ score of ≥20% in combination with an abnormal tumor MMR result. The incremental gain obtained by the addition of MSI/immunohistochemistry testing was less at higher $PREMM_{1,2}$ cut-off values; at a $PREMM_{1,2}$ cut-off of 40%, the addition of MSI/immunohistochemistry testing did not lead to an improvement in specificity.

Characteristics of Patients With Low $PREMM_{1,2}$ Scores

The $PREMM_{1,2}$ score correlated not only with the prevalence of germline mutations but also with the frequency of MMR deficiency (Table 3). Although 52 of 826 (6.3%) individuals with a $PREMM_{1,2}$ score <5% had a MSI-H tumor or showed loss of MLH1 and MSH2 on immunohistochemistry, none of them carried a germline *MLH1/MSH2* mutation and 17 (33%) were associated with *BRAF* V600E mutation in the tumor (Table 3). Interestingly, in patients with abnormal MMR tests, *BRAF* V600E mutation was significantly associated with a $PREMM_{1,2}$ score <5% (P = .009). In fact, 17 of 20 (85%) patients with a MMR-deficient tumor associated with *BRAF* V600E mutation had a $PREMM_{1,2}$ score <5%, whereas none of the 14 patients with a MMR-deficient tumor and $PREMM_{1,2}$ score ≥20% showed this variant (Table 3). As recent data demonstrate that BRAF mutations are rare in Lynch syndrome tumors, the findings are consistent with the conclusion that a low $PREMM_{1,2}$ score indicates a low likelihood that a patient with CRC has Lynch syndrome.

Characteristics of Patients With High $PREMM_{1,2}$ Scores

When patients with a $PREMM_{1,2}$ score ≥20% were stratified according to their MMR status (Table 4), patients with MMR-deficient tumors differed from those with MMR proficiency in some clinical characteristics: they are more likely women (P = .02), have a lower prevalence of previous or synchronous adenomas (P = .002), a higher prevalence of endometrial cancer (P = .003), more first-degree relatives with CRC (P = .01), and more second-degree relatives with endometrial cancer (P = .03). Therefore, a high $PREMM_{1,2}$ score in combination with MMR proficiency identified a significant group of families, recently characterized in our series, with a less penetrant cancer phenotype,[18] in line with a similar group with Amsterdam I criteria recently described as familial CRC type X syndrome.[21]

We further examined potential etiologies of the CRC in the patients with high $PREMM_{1,2}$ scores based on their MSI status. Five of 14 (36%) patients with $PREMM_{1,2}$ score ≥20% and a MMR-deficient tumor carried a *MLH1/MSH2* germline mutation. We further investigated potential etiologies of the high $PREMM_{1,2}$ scores for the 9 individuals who were not found to carry germline *MLH1/MSH2* mutations by performing supplemental analyses of PMS2 and MSH6 immunostaining, and BRAF mutation analysis. In these nonmutation carriers, no BRAF

Table 3. Prevalence of *MLH1/MSH2* Germline Mutations and Mismatch Repair Deficiency According to the $PREMM_{1,2}$ Score

$PREMM_{1,2}$ score	No.	*MLH1/MSH2* germline mutation (%)[a]	Tumor MMR deficiency[b] (%)[1]	Tumor MMR deficiency[b] associated with *BRAF* V600E mutation (%)[a]
<5%	826	– (–)	52 (6)	17 (2)
5–9%	266	2 (0.8)	18 (7)	2 (0.7)
10–19%	98	1 (1)	7 (7)	1 (1)
20–39%	23	3 (13)	7 (30)	– (–)
≥40%	9	2 (22)	7 (78)	– (–)
Total	1222	8	91	20

MMR, mismatch repair.
[a]Percentages referred to each $PREMM_{1,2}$ category.
[b]Tumor MMR deficiency demonstrated by either high microsatellite instability and/or loss of *MLH1/MSH2* expression.

Table 4. Personal and Familial Characteristics of Patients With PREMM$_{1,2}$ Score ≥20% According to the Mismatch Repair Status

	MMR proficiency (n = 18)	MMR deficiency (n = 14)	*P* value
Personal characteristics			
Age (*y*)[a]	66.0 ± 15.5	60.7 ± 19	.39
Female: no. (*%*)	4 (22.2)	9 (64.3)	.02
Previous CRC: no. (*%*)	6 (33.3)	2 (14.2)	.34
Synchronous CRC: no. (*%*)	7 (38.9)	3 (21.4)	.45
Previous or synchronous adenoma: no. (*%*)	11 (61.1)	1 (7.1)	.002
Endometrial cancer: no. (*%*)	– (–)	6 (42.9)	.003
Other Lynch-syndrome–associated cancers[b]: no. (*%*)	6 (33.3)	7 (50)	.34
Proximal CRC: no. (*%*)	7 (38.9)	6 (42.9)	.82
Familial characteristics			
Age of FDR with CRC (*y*)[a]	45.7 ± 11.3	40.4 ± 7.2	.13
FDR with CRC: no. (*%*)	6 (33.3)	11 (78.6)	.01
≥2 FDR with CRC: no. (*%*)	3 (16.7)	7 (50)	.06
FDR with endometrial cancer: no. (*%*)	7 (38.9)	3 (21.4)	.45
≥2 FDR with endometrial cancer: no. (*%*)	2 (11.1)	1 (7.1)	1.0 (%)
FDR with other Lynch-syndrome–associated cancers[b]: no. (*%*)	2 (11.1)	6 (42.9)	.09
Age of SDR with CRC (*y*)[a]	45.9 ± 7.8	46.3 ± 5.6	.84
SDR with CRC: no. (*%*)	1 (5.6)	4 (28.6)	.14
≥2 SDR with CRC: no. (*%*)	– (–)	2 (14.3)	.18
SDR with endometrial cancer: no. (*%*)	– (–)	4 (28.6)	.03
SDR with other Lynch-syndrome–associated cancers[2]: no. (*%*)	1 (5.6)	2 (14.3)	.57
Amsterdam II criteria: no. (*%*)	4 (22.2)	10 (71.4)	.005

MMR, mismatch repair; CRC, colorectal cancer; FDR, first-degree relatives; SDR, second-degree relatives.
[a]Expressed as mean ± standard deviation.
[b]Stomach, ovaries, urinary tract, small intestine, pancreas, bile ducts, brain, or sebaceous glands.

mutation was found, and normal PMS2 and MSH6 protein expression was observed in all tumors.

To better characterize the subset of patients with a PREMM$_{1,2}$ score ≥20% and MMR-proficient tumors (n = 18), MSH6 and PMS2 immunostaining and *MLH1/MSH2* germline gene testing were performed in all of them. With respect to immunostaining, normal MSH6 and PMS2 protein expression was observed in all tumors. Furthermore, *MLH1/MSH2* gene testing did not show any deleterious mutations. Finally, we performed *MYH* analysis, and one patient was found to have a biallelic *MYH* mutation (G382D/Y165C) in a nested study performed in the EPICOLON cohort.[22] The patient was a 49 year-old-man with 2 synchronous CRCs and 25 synchronous adenomas and no family history of any neoplasia.

Discussion

Extensive knowledge now available about the Lynch syndrome has encouraged researchers to look for a systematic, quantitative, and objective approach to identify these patients.[11–13,23] We recently developed the Web-based PREMM$_{1,2}$ model[11] on the basis of a logistic regression analysis from a large cohort of patients at risk for hereditary CRC who underwent genetic testing to quantify the relative importance of known clinical parameters in Lynch syndrome and predict the likelihood of carrying a mutation in the *MLH1* and *MSH2* genes. Although the model performed well among individuals at moderate risk for Lynch syndrome, its usefulness and performance in a nonselected, population-based cohort of CRC patients, either alone or in combination with tumor MMR testing, was unknown.

Our study of 1222 population-based CRC cases demonstrates that the PREMM$_{1,2}$ model constitutes a useful approach to identify *MLH1/MSH2* gene mutation carriers among patients with CRC, either alone or in combination with MMR tumor testing. The quantitative assessment of the genetic risk obtained with the PREMM$_{1,2}$ model may drive subsequent decisions about molecular testing. Moreover, the combination with tumor MMR analysis identified a sizable subgroup of patients with a heterogeneous high CRC risk, potentially involving familial CRC type X syndrome, *MYH*-related cancer, and other, still unknown inherited disorders.

The first important finding is the demonstration that a PREMM$_{1,2}$ cut-off of ≥5% identified all *MLH1* and *MSH2* mutation carriers among unselected CRC patients. The use of a higher PREMM$_{1,2}$ cut-off provided a higher specificity and PPV, at the expense of a lower sensitivity. The negative predictive value of a PREMM$_{1,2}$ score <5% was 100%, thus reinforcing the consistency of this cut-off point. Therefore, for the clinician in general practice, whose first decision point is to see if a patient with CRC

needs further molecular evaluation for Lynch Syndrome, a score of <5% indicates that no further referral is likely to be necessary, whereas a score of ≥5% should lead to further molecular evaluation. The low specificity of a 5% cut-off for the presence of germline mutations necessitates further refinement of the likelihood of carrying a mutation prior to proceeding to genetic testing. Our data demonstrate that the combination of the $PREMM_{1,2}$ score with tumor MMR testing improved its specificity and PPV. Indeed, using a cut-off of ≥5%, the addition of an abnormal tumor MMR test result provided a specificity of 98% and a PPV of 21%.

An interesting finding was that the $PREMM_{1,2}$ model, in combination with tumor MMR testing, is able to identify a subset of patients resembling the recently described familial CRC type X syndrome,[21] families who fulfill the Amsterdam I criteria without evidence of MMR deficiency. Individuals in such families have a lower incidence of CRC than those in families with Lynch syndrome, whereas incidence for other cancers may not be increased. The molecular etiology of this disorder remains unknown, with a probable heterogeneous genetic basis. In our study, patients with a $PREMM_{1,2}$ score ≥20% with MMR-proficient tumors had features similar to those with familial CRC type X syndrome[21]: weaker family history of CRC and other malignancies and lower incidence of endometrial cancer. Interestingly, one patient with a high $PREMM_{1,2}$ score and no evidence of MMR deficiency carried biallelic *MYH* mutations (Y165C/G382D).

We are aware that our study has some limitations. First, the relatively low number of patients with *MLH1/MSH2* mutations may constitute a potential drawback of the analysis, thus restraining the reliability of performance features. Second, the model does not account for *MSH6* gene mutations, although it is certain that this gene is responsible for a small proportion of Lynch syndrome cases. Finally, genetic testing was mainly performed in those patients whose tumors showed MMR deficiency, even though it is unlikely that gene carriers were undetected when both MSI analysis and immunostaining were performed systematically. In addition, to exclude this possibility, patients with a $PREMM_{1,2}$ score ≥20% also underwent genetic testing.

In the last few years, there has been much interest in establishing different strategies to improve the identification of patients with Lynch syndrome. These approaches range from using clinical criteria alone (ie, the Amsterdam criteria)[24] to universal tumor molecular testing (ie, immunostaining) in any given CRC patient.[25] The current most widely accepted recommendation, on the basis of combination of the revised Bethesda guidelines and tumor MMR testing,[7] has been found to be an effective and efficient strategy for Lynch syndrome identification.[15] However, these clinical criteria have been criticized because of the use of broad and complex variables, which make them difficult to remember for a general health care professional, their low specificity, their inability to establish the likelihood of carrying a mutation in a given patient, and the difficulty of obtaining tumor samples from affected relatives to perform the MMR analyses.[23] A potential advantage of the $PREMM_{1,2}$ model with respect to the revised Bethesda guidelines relies on its quantitative nature. In fact, the model demonstrated a reasonable ability to discriminate among risk groups for probability of mutation with respect to the prevalence of mutations observed in the *MLH1/MSH2* genes and the prevalence of MMR deficiency. The latter correlation was especially relevant given that a $PREMM_{1,2}$ score of <5% identified the subset of MMR-deficient tumors associated with the somatic *BRAF* V600E mutation, a circumstance consistent with what is seen in the sporadic CRC setting.[26–28] Taking into account these results and the performance characteristics of the predicted model's risk groups, in combination or not with tumor MMR testing, we propose a strategy for *MLH1/MSH2* genetic testing in the clinical practice (Figure 2). According to this algorithm, a $PREMM_{1,2}$ score <5% could be considered a reliable cut-off to exclude those CRC patients who do not need further risk assessment because of its 100% negative predictive value for detecting germline *MLH1/MSH2* gene mutations. In patients reaching this cut-off, further decisions could also be made on the basis of $PREMM_{1,2}$ score. In patients with a score between 5% and 19%, tumor MMR testing should be performed to achieve a reasonable PPV. Finally, considering the PPV of a $PREMM_{1,2}$ score ≥20% alone (16%) and the significant increase of the PPV at that point (Figure 1), it seems reasonable to pursue direct genetic testing in patients reaching such a score, particularly if a tumor sample is not available. It is important to note, however, that in addition to the risk estimate generated from the predictive model, other important factors (ie, accessibility to genetic services, timelines of genetic information, insurance coverage, and availability of tumor block) may help determine which strategy is the most

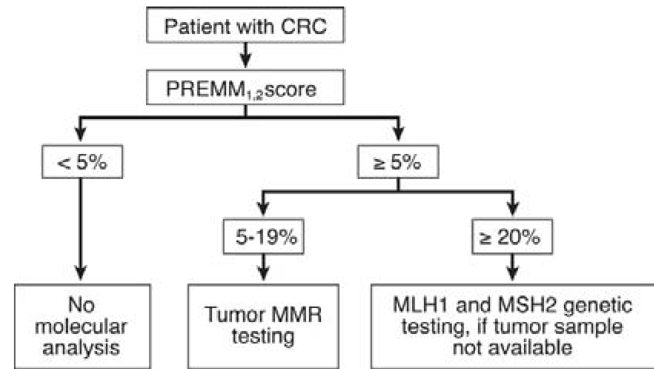


Figure 2. Proposed algorithm for the identification of *MLH1/MSH2* gene carriers among patients with colorectal cancer. CRC, colorectal cancer; MMR, mismatch repair.

convenient in a given patient. In that sense, the PREMM$_{1,2}$ model can be used by general health care providers to decide whether to refer a patient to a high-risk colorectal cancer clinic for appropriate genetic counseling, as well as by geneticists working in such units to decide on the proper molecular strategy. The use of the same algorithm in both clinical settings may contribute to a more rational referral and management approach of patients with suspected Lynch syndrome.

In conclusion, our study demonstrates that the PREMM$_{1,2}$ model is useful to identify *MLH1/MSH2* mutation carriers among unselected CRC patients. The quantitative assessment of the genetic risk might be useful to decide subsequent molecular testing and contribute to identify other high-risk individuals who may benefit from genetic risk assessment.

Appendix

Following is a list of investigators from the Gastrointestinal Oncology Group of the Spanish Gastroenterological Association who participated in the Epicolon study. Hospital 12 de Octubre, Madrid: Juan Diego Morillas (local coordinator), Raquel Muñoz, Marisa Manzano, Francisco Colina, Jose Díaz, Carolina Ibarrola, Guadalupe López, Alberto Ibáñez; Hospital Clínic, Barcelona: Antoni Castells (local coordinator), Virgínia Piñol, Sergi Castellví-Bel, Francesc Balaguer, Victòria Gonzalo, Teresa Ocaña, María Dolores Giraldez, Maria Pellisé, J. Ignasi Elizalde, Josep M. Piqué; Hospital Clínico Universitario, Zaragoza: Ángel Lanas (local coordinator), Javier Alcedo, Javier Ortego; Hospital Cristal-Piñor, Complexo Hospitalario de Ourense: Joaquin Cubiella (local coordinator), Mª Soledad Díez, Mercedes Salgado, Eloy Sánchez, Mariano Vega; Hospital del Mar, Barcelona: Montserrat Andreu (local coordinator), Xavier Bessa, Agustín Panadés, Asumpta Munné, Felipe Bory, Miguel Nieto, Agustín Seoane; Hospital Donosti, San Sebastián: Luis Bujanda (local coordinator), Juan Ignacio Arenas, Isabel Montalvo, Julio Torrado, Ángel Cosme; Hospital General Universitario de Alicante: Artemio Payá (local coordinator), Rodrigo Jover, Juan Carlos Penalva, Cristina Alenda; Hospital General de Granollers: Joaquim Rigau (local coordinator), Ángel Serrano, Anna Giménez; Hospital General de Vic: Joan Saló (local coordinator), Eduard Batiste-Alentorn, Josefina Autonell, Ramon Barniol; Hospital General Universitario de Guadalajara: Ana María García (local coordinator), Fernando Carballo, Antonio Bienvenido, Eduardo Sanz, Fernando González, Jaime Sánchez; Hospital General Universitario de Valencia: Enrique Medina (local coordinator), Jaime Cuquerella, Pilar Canelles, Miguel Martorell, José Ángel García, Francisco Quiles, Elisa Orti; Hospital do Meixoeiro, Vigo: Juan Clofent (local coordinator), Jaime Seoane, Antoni Tardío, Eugenia Sanchez; Hospital San Eloy, Baracaldo: Luis Bujanda (local coordinator), Carmen Muñoz, María del Mar Ramírez, Araceli Sánchez; Hospital Universitari Germans Trias i Pujol, Badalona: Xavier Llor (local coordinator), Rosa M. Xicola, Marta Piñol, Mercè Rosinach, Anna Roca, Elisenda Pons, José M. Hernández, Miquel A. Gassull; Hospital Universitari Mútua de Terrassa: Fernando Fernández-Bañares (local coordinator), Josep M. Viver, Antonio Salas, Jorge Espinós, Montserrat Forné, Maria Esteve; Hospital Universitari Arnau de Vilanova, Lleida: Josep M. Reñé (local coordinator), Carmen Piñol, Juan Buenestado, Joan Viñas; Hospital Universitario de Canarias: Enrique Quintero (local coordinator), David Nicolás, Adolfo Parra, Antonio Martín; Hospital Universitario La Fe, Valencia: Lidia Argüello (local coordinator), Vicente Pons, Virginia Pertejo, Teresa Sala; Hospital Universitario Reina Sofía, Córdoba: Antonio Naranjo (local coordinator), María Dolores Giraldez, María del Valle García, Patricia López, Fernando López, Rosa Ortega, Javier Briceño, Javier Padillo; Fundació Hospital Son Llatzer, Palma de Mallorca: Àngels Vilella (local coordinator), Carlos Dolz, Hernan Andreu.

References

1. Lynch HT, de la Chapelle A. Genetic susceptibility to non-polyposis colorectal cancer. J Med Genet 1999;36:801–818.
2. Giardiello FM, Brensinger JD, Petersen GM. AGA technical review on hereditary colorectal cancer and genetic testing. Gastroenterology 2001;121:198–213.
3. de la Chapelle A. The incidence of Lynch syndrome. Fam Cancer 2005;4:233–237.
4. Lynch HT, Boland CR, Gong G, et al. Phenotypic and genotypic heterogeneity in the Lynch syndrome: diagnostic, surveillance and management implications. Eur J Hum Genet 2006;14:390–402.
5. Miyaki M, Konishi M, Tanaka K, et al. Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. Nat Genet 1997;17:271–272.
6. Hendriks YM, Jagmohan-Changur S, van der Klift HM, et al. Heterozygous mutations in PMS2 cause hereditary nonpolyposis colorectal carcinoma (Lynch syndrome). Gastroenterology 2006; 130:312–322.
7. Umar A, Boland CR, Terdiman JP, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst 2004;96: 261–268.
8. Stone JG, Robertson D, Houlston RS. Immunohistochemistry for MSH2 and MHL1: a method for identifying mismatch repair deficient colorectal cancer. J Clin Pathol 2001;54:484–487.
9. Couch FJ, DeShano ML, Blackwood MA, et al. BRCA1 mutations in women attending clinics that evaluate the risk of breast cancer. N Engl J Med 1997;336:1409–1415.
10. Parmigiani G, Berry D, Aguilar O. Determining carrier probabilities for breast cancer-susceptibility genes BRCA1 and BRCA2. Am J Hum Genet 1998;62:145–158.
11. Balmana J, Stockwell DH, Steyerberg EW, et al. Prediction of MLH1 and MSH2 mutations in Lynch syndrome. JAMA 2006;296: 1469–1478.
12. Barnetson RA, Tenesa A, Farrington SM, et al. Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. N Engl J Med 2006;354:2751–2763.
13. Chen S, Wang W, Lee S, et al. Prediction of germline mutations and cancer risk in the Lynch syndrome. JAMA 2006;296:1479–1487.

14. Wijnen JT, Vasen HF, Khan PM, et al. Clinical findings with implications for genetic testing in families with clustering of colorectal cancer. N Engl J Med 1998;339:511–518.
15. Pinol V, Castells A, Andreu M, et al. Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. JAMA 2005;293:1986–1994.
16. Pinol V, Andreu M, Castells A, et al. Frequency of hereditary non-polyposis colorectal cancer and other colorectal cancer familial forms in Spain: a multicentre, prospective, nationwide study. Eur J Gastroenterol Hepatol 2004;16:39–45.
17. Xicola RM, Llor X, Pons E, et al. Performance of different microsatellite marker panels for detection of mismatch repair-deficient colorectal tumors. J Natl Cancer Inst 2007;99:244–252.
18. Llor X, Pons E, Xicola RM, et al. Differential features of colorectal cancers fulfilling Amsterdam criteria without involvement of the mutator pathway. Clin Cancer Res 2005;11:7304–7310.
19. Benlloch S, Paya A, Alenda C, et al. Detection of BRAF V600E mutation in colorectal cancer: comparison of automatic sequencing and real-time chemistry methodology. J Mol Diagn 2006;8: 540–543.
20. Sauro J, Lewis J. Estimating completion rates from small samples using binomial confidence intervals: comparisons and recommendations. Proceedings of the human factors and ergonomics society, 49th Annual Meeting 2005:2100–2104.
21. Lindor NM, Rabe K, Petersen GM, et al. Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. JAMA 2005;293:1979–1985.
22. Balaguer F, Castellvi-Bel S, Castells A, et al. Identification of MYH mutation carriers in colorectal cancer: a multicenter, case-control, population-based study. Clin Gastroenterol Hepatol 2007;5: 379–387.
23. Lipton LR, Johnson V, Cummings C, et al. Refining the Amsterdam Criteria and Bethesda Guidelines: testing algorithms for the prediction of mismatch repair mutation status in the familial cancer clinic. J Clin Oncol 2004;22:4934–4943.
24. Vasen HF, Mecklin JP, Khan PM, et al. Anticancer Res 1994;14: 1661–1664.
25. Hampel H, Frankel WL, Martin E, et al. Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). N Engl J Med 2005;352:1851–1860.
26. Weisenberger DJ, Siegmund KD, Campan M, et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nat Genet 2006;38:787–793.
27. Loughrey MB, Waring PM, Tan A, et al. Incorporation of somatic BRAF mutation testing into an algorithm for the investigation of hereditary non-polyposis colorectal cancer. Fam Cancer 2007;6: 301–310.
28. Minoo P, Moyer M, Jass J. Role of BRAF-V600E in the serrated pathway of colorectal tumourigenesis. J Pathol 2007;212:124–133.

Received July 30, 2007. Accepted October 11, 2007.

Address requests for reprints to: Antoni Castells, MD, Department of Gastroenterology, Hospital Clínic, Villarroel 170, 08036-Barcelona, Catalonia, Spain. e-mail: castells@clinic.ub.es; fax: (34) 93 227 93 87.

Francesc Balaguer and Judith Balmaña contributed equally to this work.

All authors in the Gastrointestinal Oncology Group of the Spanish Gastroenterological Association are listed in the Appendix.

Dr Llor's current address is Section of Digestive Diseases and Nutrition, University of Illinois, Chicago, Illinois.

Supported by grants from the Fondo de Investigación Sanitaria (FIS 01/0104, 03/0070, and 05/0071), the Ministerio de Educación y Ciencia (SAF 04-07190 and 07-64873), and the Asociación Española contra el Cáncer. Francesc Balaguer received a research grant from the Hospital Clínic and the Instituto de Salud Carlos III, and Sergi Castellví-Bel is supported by a contract from the Fondo de Investigación Sanitaria. Also supported by the US National Cancer Institute grant CA 113433 (Dr Syngal).

7.2. Appendix 2. "Data reduction for prediction: A case study on robust coding of age and family history for the risk of having a genetic mutation".

Ewout W.Steyerberg, Judith Balmaña, David H. Stockwell and Sapna Syngal.

Statistics in Medicine 2007 Dec 30;26(30):5545-56

STATISTICS IN MEDICINE
Statist. Med. (2007)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3119

Data reduction for prediction: A case study on robust coding of age and family history for the risk of having a genetic mutation

Ewout W. Steyerberg[1,*,†], Judith Balmaña[2,3], David H. Stockwell[4] and Sapna Syngal[2,4,5]

[1]*Department of Public Health, Erasmus MC, Rotterdam, The Netherlands*
[2]*Population Sciences Division, Dana-Farber Cancer Institute, Boston, MA, U.S.A.*
[3]*Hospital Vall d'Hebron, Universitat Autònoma de Barcelona, Spain*
[4]*Division of Gastroenterology, Brigham and Women's Hospital, Boston, MA, U.S.A.*
[5]*Harvard Medical School, Boston, MA, U.S.A.*

SUMMARY

Data reduction is often desired in the development of a prediction model, for example for effects of age and family history in the identification of subjects having a genetic mutation. We aimed to evaluate a strategy for model simplification by robust coding of related predictors. We considered 898 patients suspected of having Lynch syndrome, which is caused primarily by mutations in the mismatch repair genes, *MLH1* or *MSH2*. The presence of colorectal cancer (CRC) and endometrial cancer in patients and their relatives was related to mutation prevalence with logistic regression analysis. The performances of simplified and more complex models were quantified with a concordance statistic (c), which was corrected for optimism by cross-validation and bootstrapping. External validation was performed in 1016 patients.

The first challenge was the coding of age at diagnosis of CRC, where we forced effects to be identical in patients, in 1st degree and in 2nd degree relatives, by taking the sum of the ages at diagnosis. As a further simplification, CRC diagnosis in 2nd degree relatives was weighted half that of 1st degree relatives. These data reduction approaches were also followed for endometrial cancer. The simplified model used 7 instead of 17 degrees of freedom (df) for a more complex model incorporating individual predictor effects. The optimism-corrected c was higher (0.79 instead of 0.77), but the external c was similar (0.78 for the simplified and more complex models). A stepwise selected model performed slightly worse (external $c = 0.77$). In conclusion, a prediction model could be developed with relatively few df that captured effects of age at diagnosis across patients and relatives per type of cancer in the family. Such robust coding may especially be relevant for modeling in relatively small data sets. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: statistical modeling; Lynch syndrome; HNPCC; family history; mutation

*Correspondence to: Ewout W. Steyerberg, Department of Public Health, Erasmus MC, P.O. Box 2040, 3000 CA, Rotterdam, The Netherlands.
†E-mail: E.Steyerberg@ErasmusMC.nl

Received 24 September 2007
Accepted 24 September 2007

Copyright © 2007 John Wiley & Sons, Ltd.

1. INTRODUCTION

Lynch syndrome (or hereditary nonpolyposis colorectal cancer, HNPCC) is the most common hereditary colorectal cancer (CRC) syndrome in western countries, accounting for 2–5 per cent of all CRCs [1]. Lynch syndrome is associated with underlying mutations in the mismatch repair system, most commonly in the *MLH1* and *MSH2* genes [2]. Several guidelines have been developed to identify Lynch syndrome families, including the Amsterdam Criteria [3] and Bethesda Guidelines [4, 5]. Such guidelines intend to support health-care providers to select subjects for mutation testing, which is costly. More recently, empirically derived prediction models have been developed for the likelihood of mutations in individual patients or families, enabling a more refined selection of subjects. Some models use logistic regression [6–9], while others use Bayesian methods [10]. Aspects of family history in these models include the presence and the age at diagnosis of cancer in the proband (the index patient who is first being tested in a family), and the presence and the age at diagnosis of cancer in his/her relatives.

Modeling family history in Lynch syndrome is complex, since the spectrum of cancers associated with *MLH1* and *MSH2* mutations is diverse. Mutation carriers are mainly at risk of developing colorectal and endometrial cancer [1]. Young age at diagnosis is a risk factor of being a mutation carrier, and family members with various degrees of genetic relationship to the proband need to be considered.

We first discuss how previously developed guidelines and prediction models used age at diagnosis and family history in the prediction of the likelihood of a mutation in the proband (Section 2). These models usually focused on inclusion of statistically significant covariables. In contrast, we recently developed a model where we aimed to capture effects of age and family history with a parsimonious coding [9]. The underlying reasoning was that a slightly poorer fitting model with fewer degrees of freedom (df) would make for a more generalizable model [11].

We here aim to systematically assess alternative model specifications, including apparent, internally validated and externally validated predictive performance. Two large cohorts of individuals undergoing genetic testing were used for model development and validation (Section 3). Specific challenges arose in the analysis of age at diagnosis and the coding for the presence of 1st and 2nd degree family members with cancer (Section 4). We consider logistic models with varying complexity (Section 5) and conclude with a discussion on model complexity in predictive models (Section 6).

2. CODING IN PREVIOUS PREDICTION MODELS

The coding of age at diagnosis of cancer and family cancer history was different in previous prediction models for Lynch syndrome (Table I). Age at diagnosis was dichotomized at 50 years in recent consensus guidelines (‘Amsterdam’ [3] and ‘Bethesda’ [5], Table I). An earlier version of the Bethesda guidelines considered 45 years as a cut-off for CRC diagnosis age [4]. The model developed by Barnetson *et al*. included age of CRC diagnosis in the proband as a continuous term, but, remarkably, included cancers in relatives with a dichotomization at 50 years [8]. Other prediction models have incorporated age as a continuous predictor ([6, 7] and MMRpro [10]). Many models employ the mean age at cancer diagnosis across an entire family as a predictor of mutation status. Coding in this way implies similar age effects in probands and relatives. In other models, age has not been considered at all, or has been considered but the effects were combined

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

Table I. Overview of family history features as considered by clinical criteria and statistical models for assessment of Lynch syndrome.

Model	Age effects	Family history
Amsterdam criteria [3]	Age <50 years	At least 3 family members with CRC, endometrial, or other Lynch-associated cancers in 1st/2nd degree relatives
Revised Bethesda criteria [5]	Age <50 years for proband /1st degree relative Age not considered if # of affected relatives >2	Presence of CRC, endometrial, or other Lynch-related cancers in 1st/2nd degree relatives
Wijnen *et al.* [6]	Age <50 years for fulfillment of Amsterdam criteria Mean CRC age of all family members, including the proband No age effect for endometrial cancer	Fulfillment of the Amsterdam criteria, or # of relatives with CRC or endometrial cancer (degree of relationship not specified)
Lipton *et al.* [7]	Mean CRC or endometrial cancer age of all family members, including the proband	# of relatives with CRC, endometrial cancer, two primaries, or $\geqslant 5$ adenomas (degree of relationship with proband not specified)
Barnetson *et al.* [8]	CRC age continuous in proband CRC age <50, or $\geqslant 50$ years in relatives No age effect for endometrial cancer	Presence of CRC in relatives (degree of relationship not specified) Presence of endometrial cancer in 1st degree relatives
MMRpro [10]	CRC and endometrial cancer age continuous in proband, 1st and 2nd degree relatives Current age or age at last follow-up continuous in unaffected relatives	Mendelian distance for CRC and endometrial cancer
PREMM$_{1,2}$ [9]	CRC and adenoma age continuous in proband, CRC age in 1st, 2nd degree relatives Endometrial cancer age continuous in probands, 1st, 2nd degree relatives	Mendelian distance for CRC, endometrial, and other Lynch-associated cancers

CRC, colorectal cancer.

among different cancer types. For example, the models by Wijnen *et al.* [6] and by Barnetson *et al.* [8] considered endometrial cancer in the family, but they did not model an age effect for this diagnosis. The model by Lipton *et al.* [7] used the mean age at diagnosis of CRC and endometrial cancer, implying similar age effects for both cancers.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

Family history data were also coded in a variety of ways. Most models have used clinical information from 1st and 2nd degree relatives (Amsterdam, Bethesda, MMRpro). The types of cancers considered in relatives included CRC and endometrial cancer in all models; however, the model by Barnetson *et al*. considered endometrial cancer only in 1st degree relatives. Relatives with further genetic distance were considered in the models by Wijnen *et al*. and by Lipton *et al*. The MMRpro model explicitly uses Mendelian distance.

In summary, there are substantial differences among the different prediction models in the coding of age and family history. Issues include (1) whether age can reasonably be dichotomized at age 50; (2) whether identical age effects might be assumed for different cancers in (a) probands and relatives and (b) CRC and endometrial cancer; and (3) how genetic distance of family members of the proband should be modeled. We addressed each of these issues in the development and validation of a prediction model entitled PREMM (for Prediction of MLH1 and MSH2 Mutations).

3. PATIENTS

The patient cohorts used for model development and validation consisted of 898 and 1016 consecutive, unrelated probands from the years 2000–2003 and 2004–2005, respectively, who were undergoing genetic testing at Myriad Genetic Laboratories Inc., Salt Lake City, UT. Patients had a personal or family history suggestive of Lynch Syndrome. Testing for *MLH1* and *MSH2* mutations was ordered by health-care professionals (mainly geneticists, oncologists, gastroenterologists, or gynecologists) from across the United States [9]. Clinical data were obtained from the test order form as completed by the health-care professional ordering genetic testing, and included the patient's age, gender, and ancestry, as well as specific details about personal and family cancer history. Family history of these probands included colorectal and endometrial cancer, as well as other Lynch Syndrome-associated cancers. The test order form asked specifically for maternal or paternal origin of each relative. In one instance, both sides of the family were affected by Lynch-associated cancers, and the lineage most likely to be affected was included.

Laboratory methods for full gene sequencing have been described in detail before [9]. Genetic variants were detected by comparison with a consensus wild-type sequence constructed for each gene. We classified subjects found to have either deleterious or suspected deleterious mutations as 'mutation positive.' Subjects with all other genetic variants and polymorphisms were included in the 'mutation negative' group.

4. MODELING AGE AT DIAGNOSIS AND FAMILY HISTORY

We used logistic regression analysis to relate proband and family history characteristics to the presence of a mutation. An indicator variable was created for probands having multiple CRCs since this group was reasonably large. For these, age at diagnosis was defined at the first CRC diagnosis. This definition of age was considered more practical than using the mean age of diagnoses, which had a similar regression coefficient. For relatives with CRC, we created variables indicating both the number of affected relatives (1 *versus* 2 or more) and their relationship to the proband (1st *versus* 2nd degree). We included two variables for each diagnosis in the multivariable model: an indicator variable for the presence or the absence of that diagnosis and a variable relating to the

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

age at diagnosis. This approach was followed for proband diagnoses and relative diagnoses and repeated for endometrial cancer diagnoses in the family.

Statistical evaluation included Akaike's information criterion (AIC) as a measure of model fit with correction for df used in the model (difference in -2 loglikelihood $-2\,\mathrm{df}$), and a concordance statistic (c) to indicate discrimination between patients with and without mutations [12]. The c statistic is similar to the area under the receiver operating characteristic curve for binary outcomes. It was first determined by evaluation of predictions on the development sample (apparent performance). Next, internal validation was performed [13]. Model selection steps are difficult to take into account in AIC but were considered in a cross-validation and bootstrap resampling procedure [11]. For cross-validation, the data set was randomly divided into five equally sized parts. Models were developed in four of the parts, including model selection, and validated in the remaining fifth part. The average performance was taken over the five validation parts that were left out once each, and the whole procedure was repeated 10 times to increase stability (10 times 5-fold cross-validation). For the bootstrap procedure, 200 random samples were drawn with replacement, and predictive models were developed in each sample (including any model selection step) and evaluated in the development cohort. External validation was done in the 1016 patients in the validation cohort. We used the Design library in S-Plus V6 software (Insightful, Inc., Seattle, WA).

5. RESULTS

5.1. Univariate analyses

The prevalence of mutations was 14 per cent (130/898) in the development cohort and similar in the validation cohort (15 per cent, 155/1016, Table II). In the development cohort, strong predictive effects were noted for CRC and endometrial cancer, both in the proband and in 1st degree relatives, but not in 2nd degree relatives. If multiple CRCs occurred in the same patient, it was especially predictive of an underlying genetic cause (prevalence of mutations 18/40, 45 per cent, Table II).

The age at diagnosis in probands ranged between 9 and 89 years for CRC (interquartile range: 37–51 years) and between 20 and 72 years for women with endometrial cancer (interquartile range: 39–54 years). CRC and endometrial cancer diagnoses were on average 5 and 2 years earlier in those with mutations than in those without mutations, respectively.

5.2. A model for CRC family history

On the basis of the previous models as shown in Table I, we formulated a 'full' logistic regression model for the effects of CRC in probands and relatives. We consider the main effects for probands with 1 or multiple CRC diagnoses ('CRC1' and 'CRC2'); 1st degree relatives (1 affected relative or 2 or more affected relatives, 'CRC1st1' and 'CRC1st2'); 2nd degree relatives (1 or 2+, 'CRC2nd1' and 'CRC2nd2'); and 3 age effects (separately for probands, 1st degree, and 2nd degree relatives). Age was scaled by the factor 10 at its approximate mean at diagnosis of 45 years.

Multivariable odds ratios (ORs) were readily interpretable and in line with the univariate results (Table III). Age effects were rather similar, with an OR of approximately 0.6 per 10-year younger diagnosis. This model had 9 df (Table IV). The model likelihood ratio (LR) was 100.4. AIC was $100.4-2*9=82.4$, with apparent c statistic$\,=0.761$. Cross-validation and bootstrapping both indicated a decrease in performance for future patients ($c=0.738$ and 0.745, respectively). External validation confirmed the anticipated discriminative ability ($c=0.728$).

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

Table II. Characteristics and family history of probands in the development and validation cohorts according to mutation status.

	Prevalence of mutations	
Characteristic	Development 14 per cent (130/898)	Validation 15 per cent (155/1016)
Proband		
CRC		
0	11 per cent (40/362)	8 per cent (39/466)
1 CRC	15 per cent (72/496)	19 per cent (98/509)
2 or more CRC	45 per cent (18/40)	44 per cent (18/41)
Adenoma	15 per cent (21/141)	10 per cent (24/231)
Endometrial cancer	24 per cent (22/91)	30 per cent (45/149)
Other Lynch-associated cancer	22 per cent (22/100)	19 per cent (18/93)
1st degree family history		
Relatives with CRC		
0	9 per cent (37/396)	10 per cent (50/491)
1	16 per cent (59/363)	17 per cent (62/372)
2 or more	24 per cent (34/139)	28 per cent (43/153)
Relatives with endometrial cancer		
0	13 per cent (101/804)	14 per cent (126/902)
1	28 per cent (24/85)	22 per cent (23/105)
2 or more	56 per cent (5/9)	67 per cent (6/9)
2nd degree family history		
Relatives with CRC		
0	15 per cent (81/528)	16 per cent (102/651)
1	13 per cent (32/241)	14 per cent (33/229)
2 or more	13 per cent (17/129)	15 per cent (20/136)
Relatives with endometrial cancer		
0	14 per cent (119/851)	15 per cent (137/939)
1	24 per cent (10/41)	22 per cent (15/69)
2 or more	17 per cent (1/6)	38 per cent (3/8)

CRC, colorectal cancer.

5.3. *Categorization and non-linearity in age effects*

As expected, dichotomization of age at 50 years led to a poorer performance than using linear, continuous effects of age. At internal and external validation, the c statistic of a model with dichotomized age was 0.64 rather than 0.74 and 0.69 rather than 0.73, respectively.

We explored non-linearity in age effects in three ways: with polynomials ($x+x^2$ and $x+x^2+x^3$), restricted cubic spline functions (3 knots, 2 df, and 4 knots, 3 df) [14], and a breakpoint model, where age at diagnosis had an effect after age 50 only. Some non-linearity was found for age at CRC diagnosis in 1st degree relatives (x^2 in model with $x+x^2$: $p=0.02$, spline term in model with 3 knots: $p=0.04$). Extending the model with a square term ($x+x^2$) led to similar performance at internal validation but a slightly poorer external validity (c 0.723 instead of 0.728). A breakpoint model led to poorer performance than a model with linear terms. In sum, using linear terms for age effects was better than dichotomization at age 50 or extension with non-linear terms.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

Table III. Multivariable odds ratios of CRC and age of CRC diagnosis for mutations in the development cohort ($n=898$, 130 mutations).

	CRC full model	CRC bw AIC	CRC parsimonious age+CRCfam
Proband			
CRC			
0	1	1	1
1 CRC	1.4 [0.9–2.2]	1.3 [0.9–2.1]	1.4 [0.9–2.1]
2 or more CRC	6.6 [3.1–14.3]	6.0 [2.8–12.7]	6.2 [2.9–13.3]
Age (per decade)	0.67 [0.54–0.84]	0.65 [0.52–0.82]	0.66 [0.59–0.75]*
1st degree family history			
Relatives with CRC			1.7 [1.3–2.2]**
0	1	1	
1	2.4 [1.5–3.8]	2.5 [1.6–3.9]	
2 or more	3.0 [1.7–5.4]	3.2 [1.8–5.7]	
Age (per decade)	0.68 [0.55–0.84]	0.66 [0.53–0.81]	0.66 [0.59–0.75]*
2nd degree family history			
Relatives with CRC		—	1.3 [1.1–1.5]**
0	1		
1	1.4 [0.8–2.3]		
2 or more	0.8 [0.4–1.5]		
Age (per decade)	0.59 [0.45–0.77]	—	0.66 [0.59–0.75]*

CRC, colorectal cancer; bw AIC, backward stepwise selection with Akaike's information criterion.
*Age effect forced to be identical for proband, 1st, and 2nd degree relatives.
**Family history effect based on assuming 50 per cent effect in 2nd degree *versus* 1st degree relatives, and linearity of effect for 2 *versus* 1 *versus* no 1st and 2nd degree relatives, respectively.

5.4. Stepwise selection in CRC model

One option for model simplification is backward stepwise selection. We used the AIC criterion as the stopping rule, which is identical to requiring a p-value less than 0.157 for variables with 1 df in the model. The 2nd degree CRC effects were removed, leading to a model with 6 rather than 10 df (Table III). Predictive performance was slightly less than the original model at internal and external validation (Table IV).

5.5. Similarity in CRC age effects

As an alternative model simplification, we created a summary variable 'sumCRCage' for the 3 age variables ('CRCage' for CRC diagnosis in the proband; 'CRCage1st' and 'CRCage2nd' for 1st and 2nd degree relatives, respectively): sumCRCage = CRCage + CRCage1st + CRCage2nd. The 3 age coefficients were forced to be the same by using the sum of the ages at diagnosis rather than the 3 age variables separately in a logistic regression model. This led to 7 instead of 9 df. The age effect was an OR of 0.65 [0.58–0.74] per decade. Model performance was similar to the original model (Table IV).

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

Table IV. Performance of alternative models including predictive effects of CRC and endometrial cancer and age of diagnosis for mutations.

Cancers considered	Model	Terms	Df	AIC	C statistic: Apparent	C statistic: Cross-validated	C statistic: Boot	C statistic: External validation
CRC	Full	CRC1+CRC2+CRCage+ CRC1st1+CRC1st2+ CRC2nd1+CRC2nd2+ CRC1stAge+CRC2ndAge	9	82.4	0.761	0.738	0.745	0.728
	Age dichotomized	CRC1 < 50+CRC2 < 50 +CRC1st1<50+CRC1st2<50 +CRC2nd1 < 50+CRC2nd2 < 50	6	32.3	0.660	0.640	0.643	0.689
	bw AIC	CRC1+CRC2+CRCage+ CRC1st1+CRC1st2+ CRC1stAge	6	70.3	0.733	0.726	0.728	0.715
	Parsimonious age	CRC1+CRC2+CRC1st1+ CRC1st2+CRC2nd1+ CRC2nd2+sumCRCage	7	85.7	0.761	0.743	0.748	0.728
	Parsimonious age+ CRCfam	CRC1+CRC2+CRCfam+ sumCRCage	4	86.2	0.752	0.744	0.748	0.725
CRC+ endometrial cancer	Full	CRC1+CRC2+CRCage+ CRC1st1+CRC1st2+ CRC2nd1+CRC2nd2+ CRC1stAge+CRC2ndAge+ Endo+EndoAge+Endo1st1+ Endo1st2+Endo2nd1+ Endo2nd2+Endo1stAge+ Endo2ndAge	17	93.1	0.800	0.753	0.769	0.782
	Simplified	CRC1+CRC2+CRCfam+ sumCRCage+Endo+ Endofam+sumEndoAge	7	110.9	0.797	0.784	0.786	0.782
	bw AIC	CRC1+CRC2+CRCage+ CRC1st1+CRC1st2+ CRC1stAge+Endo+ Endo1st1+Endo1st2+ Endo2nd1	10	92.6	0.796	0.755	0.768	0.765

CRC, colorectal cancer; Endo, endometrial cancer; bw AIC, backward stepwise selection with AIC for selection of model terms; cross-validation, 10 times 5-fold cross-validation; boot, bootstrap validation with 200 replications; external validation, assessment of performance in 1018 independent patients (see text). CRC1, CRC in the proband; CRC2, multiple CRCs in the proband; CRCage, age of CRC in the proband; CRC1st1, CRC1st2, 1 or 2 CRC in 1st degree relatives; CRC2nd1, CRC2nd2, 1 or 2 CRC in 2nd degree relatives; CRC1stAge, CRC2ndAge, minimum age of CRC diagnosis in 1st and 2nd degree relatives; sumCRCage, CRCage+CRC1stAge+CRC2ndAge; CRCfam, CRC1st1+CRC1st2+0.5∗CRC2nd1+0.5∗CRC2nd2.

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

5.6. *Coding CRC family history*

We combined the family history of 1st and 2nd degree relatives as CRCfam = CRC1st + 0.5 ∗ CRC2nd, with CRC1st referring to the number of 1st degree relatives (0, 1, 2+) and CRC2nd the number of 2nd degree relatives (0, 1, 2+). The CRCfam variable can be used instead of 4 indicator variables for having 1st and 2nd degree relatives with CRC. Hence, we gain 3 df by using the Mendelian genetic distance between 1st and 2nd degree relatives (50 per cent), and assuming that 2 or more *versus* 1 affected relative increases the odds of having a mutation by a factor 2 (linearity in effect of number of relatives). The effects of age and family history are still readily interpretable as shown in Table III. The performance of this model with 4 df was slightly better than the original model according to AIC and internal validation procedures but slightly worse at external validation. Apparently, the benefit of saving 3 df was less than the price of making simplifying assumptions for the coding in a single summary variable (‘CRCfam’).

5.7. *Effects of CRC and endometrial cancer*

Next, a prediction model was built with endometrial cancer diagnoses in addition to CRC diagnoses. We created an indicator variable whose reference category was females without endometrial cancer and all males (‘Endo’, coded 0/1). Age effects were considered separately for probands, 1st degree, and 2nd degree relatives. This model had 17 parameters. Both AIC and internally validated c statistics were larger compared with models with CRC only. External validation showed a clear increase in c statistic (from 0.73 to 0.78, Table IV).

For simplification, age at diagnosis in the proband was combined with age at diagnosis in 1st and 2nd degree relatives, and family history was coded as for CRC: number of affected 1st degree relatives $(0,1,2+)+0.5$ times number of affected 2nd degree relatives $(0,1,2+)$. These simplifications led to better internally validated performance than the original more complex model (Table IV). External validation revealed a similar performance as the original model ($c=0.78$).

As a further simplification we estimated identical age effects for CRC and endometrial cancer diagnoses by summing all six original age variables. This saved another df (6 instead of 7 df) and led to slightly better AIC (112), as well as better internal and external validity. The external c statistic was 0.785.

For comparison, we again applied backward stepwise selection with AIC to simplify the full model. The 2nd degree CRC effects were removed, as well as all endometrial cancer age effects. The latter age effects had p-values of 0.88, 0.45, and 0.36 in the full model; when the three effects were combined, the p-value was 0.18 in the simplified model. The backward selected model had 10 instead of 17 df. Predictive performance was slightly less than the full or simplified model at internal and external validation (Table IV).

6. DISCUSSION

This study shows that effects of age and family history can be captured with a limited number of summary variables to predict the presence of an underlying mutation. This approach may especially be valuable in genetic epidemiologic studies with relatively few mutations, when we aim to develop a prediction model for the probability of carrying a mutation. Parsimony is less important with larger sample sizes. The effect of age at diagnosis was reasonably approximated with a single linear term for probands and relatives. The estimation of one single parameter for

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

the age effect in probands, 1st degree, and 2nd degree relatives led to a better optimism-corrected performance than estimating separate age effects. Further, the genetic distance between family members could easily be used to derive a summary variable. This second simplification led to a similar optimism-corrected performance as allowing for separate effects of having 1 or 2 affected 1st or 2nd degree relatives. External validation did not support the latter simplification.

The results of internal and external validation were not fully in agreement. Possibly not all modeling steps were represented in the comparison of simplified models to full models. Also, the external validation comprised a limited number of events (155 mutations), which leaves some uncertainty in the results. The value of simplified coding of family history of CRC, endometrial cancer, and other Lynch-associated cancers in the PREMM model hence needs further external validation [9]. It will especially be of interest to compare PREMM with other recent prediction models [15], which relied on stepwise selection for model specification. We might expect a worse performance for the Barnetson *et al*. model, since it was based on analysis of only 38 mutations, with dichotomization of CRC age, absence of endometrial cancer age, and a less comprehensive spectrum of Lynch-associated tumors in the family history [8].

The proposed coding exploits inherent relationships between predictor variables in an attempt to limit overfitting. Overfitting has been described as 'the curse of prediction modeling' [16]. On the other hand, we should not oversimplify predictor–outcome relationships. The combination of subject matter knowledge and empirical data should guide us in finding an appropriate balance between overfitting and oversimplification. In our case study, several clinical guidelines and statistical models had already been developed, suggesting which predictors should be included in the model. For example, CRC and endometrial cancer are the core parts of the spectrum of Lynch syndrome. Also, it was clear that cancer diagnoses in 1st and 2nd degree relatives should be considered in addition to information of the proband. The challenge was merely in the optimal coding of these predictors *versus* trying to estimate a relatively complex model with individual effects of many covariates.

One aspect was the coding of continuous predictors, especially age at diagnosis. We found that a linear relationship was a quite reasonable assumption and led to a better performance than dichotomizing age at 50 years, as was done in many other models. Some clinicians may like the simplicity of an age cut-off. It is however well known that dichotomization implies a loss of information [17].

To further limit overfitting, we forced the age effects of similar phenomena to be identical, e.g. age at diagnosis of CRC. This approach was implicitly followed by many other models, which included a mean age at diagnosis over probands and relatives. Our explicit approach allows for formal testing, since the more complex and simplified models can be formulated as nested models. For example, a model with a common age effect for CRC (based on three ages) could be extended with two separate age effects for CRC age in 1st degree and 2nd degree relatives, providing the basis for an LR, Wald, or Score test with 2 df.

For estimation, we might also have considered a penalized maximum likelihood approach, with penalty on deviations from a common age effect [12, 18]. A simple heuristic shrinkage approach could also be envisioned, where we use the LR statistic of a model extension to shrink the regression coefficients of that extension, accounting for the df used for the extension (shrinkage factor $=(\mathrm{LR}-\mathrm{df})/\mathrm{LR}$) [19].

We expected that CRC and endometrial cancer diagnosis age would have different effects and hence would model these separately, in contrast to the model by Lipton *et al*. [7] (Table I). Unexpectedly, we found that assuming a single age effect for CRC and endometrial cancer actually

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

led to a better prediction model. The models by Wijnen *et al*. and Barnetson *et al*. did not include any effect for age at diagnosis of endometrial cancer, which likely reflects a lack of power [6, 8]. The model by Barnetson *et al*. had only 38 mutations for model development, and endometrial cancer in family members was infrequent. In our development data set, the age effects of endometrial cancer were not statistically significant.

Finally, the incorporation of genetic distance and type of cancer was addressed differently in the previous models. Most logistic regression models simply considered affected relatives irrespective of whether the cancer was CRC, endometrial cancer, or other Lynch-associated cancer (Table I). Our approach was to weight cancer diagnoses in 2nd degree relatives as half that of 1st degree relatives and use separate coefficients for each cancer. The most detailed weighting was performed in the MMRpro model, which also takes the age of unaffected relatives into account in a Bayesian calculation [10]. This is similar to the idea of using the observed minus expected cases to quantify the family history [20]. A recent study lists a number of other family history scores [21].

A number of limitations in our study should be mentioned. Our cohorts consisted of subjects considered at risk of Lynch syndrome according to their caregivers; it was not a truly population-based study. We relied on information from the test ordering form, as stated by the health-care provider. This may have been unreliable for some cases and was missing in almost a quarter of the patients initially considered in the development cohort. Self-reported family history may be even more unreliable [22]. We did not consider age and number of unaffected relatives, nor did we analyze sib-pairs and parent–offspring pairs separately. We did not have data to model testing for microsatellite instability, while this is important in the diagnostic workup for Lynch syndrome [23]. Also, information was not available on MSH6 mutations, which also cause Lynch syndrome. Finally, one might see the Lynch-associated cancers as a competing risk problem, which would lead to another modeling approach than logistic regression.

In conclusion, family history could reasonably be summarized by a straightforward weighing of 2nd *versus* 1st degree relatives, and assuming a single age effect across patients and relatives per type of cancer diagnosis. This approach may be valuable in other prediction models to robustly capture most of the complex information from family history.

ACKNOWLEDGEMENTS

The authors like to acknowledge Myriad Genetic Laboratories Inc. for their collaboration and for providing the data for this analysis.

REFERENCES

1. Lynch HT, de la Chapelle A. Genetic susceptibility to non-polyposis colorectal cancer. *Journal of Medical Genetics* 1999; **36**(11):801–818.
2. Chung DC, Rustgi AK. The hereditary nonpolyposis colorectal cancer syndrome: genetics and clinical implications. *Annals of Internal Medicine* 2003; **138**(7):560–570.
3. Bellacosa A, Genuardi M, Anti M, Viel A, Ponz de Leon M. Hereditary nonpolyposis colorectal cancer: review of clinical, molecular genetics, and counseling aspects. *American Journal of Medical Genetics* 1996; **62**(4):353–364.
4. Rodriguez-Bigas MA, Boland CR, Hamilton SR, Henson DE, Jass JR, Khan PM, Lynch H, Perucho M, Smyrk T, Sobin L, Srivastava S. A National Cancer Institute workshop on hereditary nonpolyposis colorectal cancer syndrome: meeting highlights and Bethesda guidelines. *Journal of the National Cancer Institute* 1997; **89**(23):1758–1762.
5. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R, Hamilton SR, Hiatt RA, Jass J, Lindblom A, Lynch HT, Peltomaki P, Ramsey SD, Rodriguez-Bigas MA, Vasen HF, Hawk ET, Barrett JC, Freedman AN, Srivastava S. Revised Bethesda guidelines for hereditary

Copyright © 2007 John Wiley & Sons, Ltd.

DOI: 10.1002/sim

nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *Journal of the National Cancer Institute* 2004; **96**(4):261–268.

6. Wijnen JT, Vasen HF, Khan PM, Zwinderman AH, van der Klift H, Mulder A, Tops C, Moller P, Fodde R. Clinical findings with implications for genetic testing in families with clustering of colorectal cancer. *New England Journal of Medicine* 1998; **339**(8):511–518.
7. Lipton LR, Johnson V, Cummings C, Fisher S, Risby P, Eftekhar Sadat AT, Cranston T, Izatt L, Sasieni P, Hodgson SV, Thomas HJ, Tomlinson IP. Refining the Amsterdam criteria and Bethesda guidelines: testing algorithms for the prediction of mismatch repair mutation status in the familial cancer clinic. *Journal of Clinical Oncology* 2004; **22**(24):4934–4943.
8. Barnetson RA, Tenesa A, Farrington SM, Nicholl ID, Cetnarskyj R, Porteous ME, Campbell H, Dunlop MG. Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. *New England Journal of Medicine* 2006; **354**(26):2751–2763.
9. Balmana J, Stockwell DH, Steyerberg EW, Stoffel EM, Deffenbaugh AM, Reid JE, Ward B, Scholl T, Hendrickson B, Tazelaar J, Burbidge LA, Syngal S. Prediction of MLH1 and MSH2 mutations in Lynch syndrome. *Journal of the American Medical Association* 2006; **296**(12):1469–1478.
10. Chen S, Wang W, Lee S, Nafa K, Lee J, Romans K, Watson P, Gruber SB, Euhus D, Kinzler KW, Jass J, Gallinger S, Lindor NM, Casey G, Ellis N, Giardiello FM, Offit K, Parmigiani G. Prediction of germline mutations and cancer risk in the Lynch syndrome. *Journal of the American Medical Association* 2006; **296**(12):1479–1487.
11. Hastie T, Tibshirani R, Friedman JH. *The Elements of Statistical Learning*: *Data Mining*, *Inference*, *and Prediction*. Springer: New York, 2001.
12. Harrell FE. *Regression Modeling Strategies*: *With Applications to Linear Models*, *Logistic Regression*, *and Survival Analysis*. Springer: New York, 2001.
13. Steyerberg EW, Harrell Jr FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology* 2001; **54**(8):774–781.
14. Harrell Jr FE, Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. *Journal of the National Cancer Institute* 1988; **80**(15):1198–1202.
15. Ford JM, Whittemore AS. Predicting and preventing hereditary colorectal cancer. *Journal of the American Medical Association* 2006; **296**(12):1521–1523.
16. van Houwelingen HC. Validation, calibration, revision and combination of prognostic survival models. *Statistics in Medicine* 2000; **19**(24):3401–3415.
17. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Statistics in Medicine* 2006; **25**(1):127–141.
18. Verweij PJ, Van Houwelingen HC. Penalized likelihood in Cox regression. *Statistics in Medicine* 1994; **13**(23–24):2427–2436.
19. Steyerberg EW, Borsboom GJ, Van Houwelingen HC, Eijkemans MJ, Habbema JD. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Statistics in Medicine* 2004; **23**(16):2567–2586.
20. Houwing-Duistermaat JJ, Van Houwelingen HC. Incorporation of family history in logistic regression models. *Statistics in Medicine* 1998; **17**(24):2865–2882.
21. Murad H, Kalter-Leibovici O, Chetrit A, Freedman LS. A statistical comparison of different family history scores. *Statistics in Medicine* 2007; **26**(14):2785–2798.
22. Ramsey SD, Burke W, Pinsky L, Clarke L, Newcomb P, Khoury MJ. Family history assessment to detect increased risk for colorectal cancer: conceptual considerations and a preliminary economic analysis. *Cancer Epidemiology Biomarkers and Prevention* 2005; **11**:2494–2500.
23. Loukola A, de la Chapelle A, Aaltonen LA. Strategies for screening for hereditary non-polyposis colorectal cancer. *Journal of Medical Genetics* 1999; **36**(11):819–822.

Copyright © 2007 John Wiley & Sons, Ltd.

7.3. Appendix 3. "Phenotype comparison of MLH1 and MSH2 mutation carriers in a cohort of 1,941 individuals undergoing clinical genetic testing in the United States".

Fay Kastrinos, Elena M.Stoffel, Judith Balmaña, Ewout W. Steyerberg, Rowena Mercado, Sapna Syngal.

Cancer Epidemiol Biomarkers Prev 2008;17 (8):2044-2051

Phenotype Comparison of *MLH1* and *MSH2* Mutation Carriers in a Cohort of 1,914 Individuals Undergoing Clinical Genetic Testing in the United States

Fay Kastrinos,[1,3] Elena M. Stoffel,[1,2,3] Judith Balmaña,[4] Ewout W. Steyerberg,[5] Rowena Mercado,[2] and Sapna Syngal[1,2,3]

[1]Division of Gastroenterology, Brigham and Women's Hospital; [2]Population Sciences Division, Dana-Farber Cancer Institute; and [3]Harvard Medical School, Boston, Massachusetts; [4]Medical Oncology Department, Hospital Vall d'Hebrón, Universitat Autònoma de Barcelona, Barcelona, Spain; and [5]Department of Public Health, Erasmus Medical Center, Rotterdam, the Netherlands

Abstract

Background and Aims: **Lynch syndrome is caused by germ-line mismatch repair gene mutations. We examined the phenotypic differences between *MLH1* and *MSH2* gene mutation carriers and whether mutation type (point versus large rearrangement) affected phenotypic expression.**
Methods: **This is a cross-sectional prevalence study of 1,914 unrelated probands undergoing clinical genetic testing for *MLH1* and *MSH2* mutations at a commercial laboratory.**
Results: **Fifteen percent (285 of 1,914) of subjects had pathogenic mutations (112 *MLH1*, 173 *MSH2*). *MLH1* carriers had a higher prevalence of colorectal cancer (79% versus 69%, $P = 0.08$) and younger mean age at diagnosis (42.2 versus 44.8 years, $P = 0.03$) than *MSH2* carriers. Forty-one percent of female carriers had endometrial cancer and prevalence was similar in both groups. Other cancers were more frequent in *MSH2* carriers (24% versus 9%, $P = 0.001$) and their families ($P < 0.001$). Multivariable analyses confirmed these associations. Of the 1,016 subjects who underwent Southern blot analysis, 42 had large rearrangements (7 *MLH1*, 35 *MSH2*). There were no phenotypic differences between carriers with large rearrangements and point mutations.**
Conclusions: **In this large study of mismatch repair gene mutation carriers from the United States, *MLH1* carriers had more colorectal cancer than *MSH2* carriers whereas endometrial cancer prevalence was similar. Large genomic rearrangements were more frequent in the *MSH2* gene. *MSH2* carriers and their relatives have more extracolonic nonendometrial Lynch syndrome–associated cancers and may benefit from additional screening.** (Cancer Epidemiol Biomarkers Prev 2008;17(8):2044–51)

Introduction

Lynch syndrome, also known as hereditary nonpolyposis colorectal cancer (HNPCC), is the most common inherited colorectal cancer syndrome. It is estimated to account for up to 5% of all colorectal cancers and is caused predominantly by a mutation in one of four DNA mismatch repair genes: *MSH2, MLH1, MSH6*, and *PMS2*. Germ-line alterations in the associated cancer susceptibility genes seem to confer a 60% to 80% lifetime risk of developing colorectal cancer in the absence of medical intervention (1-4).

Mismatch repair gene mutations are also associated with a significantly increased risk for certain types of extracolonic malignancies. Lifetime endometrial cancer risk is estimated to be 40% to 60%, and risks for ovarian cancer and tumors of the urinary bladder and renal collecting system are believed to range from 10% to 20% (5-7). Other gastrointestinal tumors such as stomach, hepatobiliary, and pancreatic carcinomas, although overrepresented in Lynch syndrome compared with their prevalence in the general population, seem to occur less commonly. The possibility that cancer risks may vary depending on the type of mismatch repair gene mutation may have significant implications on cancer screening recommendations.

Despite the major advances made in molecular genetics, there remains a limited awareness of the syndrome in the general medical community. Deficiencies in family information and incomplete gene penetrance have contributed to the underrecognition of affected patients and families. Clinicians are challenged by the heterogeneity that exists when assessing the Lynch syndrome phenotype (8). Multiple predictive models have been recently developed to aid in the recognition of mutation carriers (9-11). However, whether there is a difference in genotype-phenotype expression between carriers of the two most common mismatch repair gene mutations, *MLH1* and *MSH2*, has not been resolved. The existing data are limited and derived predominantly from European family registries that invariably are subject to ascertainment bias.

In the present study, we examine the genotype-phenotype differences in Lynch syndrome mutation carriers in a large United States population undergoing clinical genetic testing for mutations in the *MLH1* and *MSH2* genes. Because different mutation types (point

Received 4/3/08; revised 5/27/08; accepted 6/5/08.

Grant support: National Cancer Institute grant K24 CA113433 (S. Syngal). National Cancer Institute grant K07 CA120448-02 (E. Stoffel).

Requests for reprints: Sapna Syngal, Dana-Farber Cancer Institute, 1 Jimmy Fund Way, SM 209, Boston, MA 02115. Phone: 617-632-5022; Fax: 617-632-4088. E-mail: ssyngal@partners.org

Copyright © 2008 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-08-0301

mutations versus large genomic rearrangements) may have different functional consequences on *MLH1* and *MSH2* genes, correlations between mutation type and phenotypic expression were also explored.

Materials and Methods

Study Population. A total of 2,276 unrelated probands submitted blood samples for full gene sequencing of *MLH1* and *MSH2* to Myriad Genetic Laboratories, Inc. (Salt Lake City, UT), starting in 2000. Information on the proband demographic profile, including age, gender, and ethnicity, as well as personal and family cancer history was obtained from the test order form that was completed by the health care professional ordering the genetic testing. Because of missing information, 362 probands were excluded. Data from 1,914 probands were available for analysis (9).

The total number of relatives per proband included those who were (*a*) first-degree relatives (FDR) or second-degree relatives (SDR) of the probands; (*b*) affected with the following Lynch syndrome–related cancers: colorectum, endometrium, stomach, small intestine, pancreas, bile ducts, ovaries, urinary tract (kidney, ureter, bladder), brain (glioblastoma multiforme), or sebaceous glands; and (*c*) on the affected side of the family. Overall, there were a total of 4,361 relatives with Lynch syndrome–related cancers in the final study sample.

Data collection and analysis occurred independently. Collection of clinical information and molecular analyses occurred at Myriad Genetic Laboratories and an anonymized, electronic data set was provided to investigators at Dana-Farber/Harvard Cancer Center for further statistical analyses. This study was reviewed and approved by the Dana-Farber/Harvard Cancer Center Institutional Review Board.

Laboratory Methods. Mutation analysis was done using methods previously described (9). Full gene sequencing of *MLH1* and *MSH2* was done on 898 unrelated probands that submitted blood samples to Myriad Genetics starting in 2000. Starting in August 2004, full gene sequencing and large genomic rearrangement analysis of *MLH1* and *MSH2* genes were done on 1,016 additional blood samples from unrelated probands. Southern blot analysis was used to search for large genomic alterations in the *MLH1* and *MSH2* genes. Methylation analyses to evaluate the presence of *MLH1* or *MSH2* germ-line epimutations were not done on the samples provided by the subjects in this study.

Individuals with deleterious mutations or "suspected deleterious" mutations were defined as mutation positive. Mutations leading to truncating or unstable proteins were considered deleterious and included frameshift, nonsense, and splice site mutations, as well as large deletions and rearrangements. All exon deletions and nonsense and frameshift mutations occurring at or before amino acids 733 and 888 of *MLH1* and *MSH2*, respectively, were defined as mutation-positive in this study. "Suspected deleterious" mutations were those genetic variants for which the available evidence indicated a likelihood (but no proof) that the mutation was deleterious. Examples include splice site mutations that occur at the conserved locations of splice acceptors and splice donors. In this study, missense mutations and noncoding intervening sequence mutations were defined as deleterious based on data derived from linkage analysis of high-risk families, functional assays, biochemical evidence, and/or demonstration of abnormal mRNA transcript processing. Mutation-negative probands were individuals with identified missense mutations within intronic regions whose clinical significance is not yet established, as well as those with polymorphisms and unclassified variants.

Statistical Methods. Statistical analysis was done using SAS statistical software (version 9.1; SAS Institute, Inc.). The variables related to the proband included presence and age(s) of colorectal cancer (none, one, two, or more), endometrial cancer, and/or other Lynch syndrome–related cancer [stomach, small intestine, pancreas, bile ducts, ovaries, urinary tract (kidney, ureter, bladder), brain (glioblastoma multiforme), or sebaceous glands]. Age was treated as a continuous variable and thereafter dichotomized to <50 or ≥50 years. The variables related to the proband family members included number of colorectal cancers, endometrial cancers, and/or other Lynch syndrome–related cancers and corresponding age(s) of diagnosis, as well as relationship to proband (FDR or SDR). To incorporate the effect of genetic distance between FDR and SDR, a weighted sum of cancer diagnoses among relatives was used. Cancer diagnoses in SDR were weighted half that in FDR (12). In probands and family members diagnosed as having the same cancer more than once, the age at diagnosis was defined as the youngest age.

Univariate analyses were used to assess the relationship between mismatch repair genes mutation status or mutation type and potential predictive variables related to the carriers' personal and family history of cancer. Categorical outcome data were reported as frequencies. Comparisons between mismatch repair mutation status groups (*MLH1* versus *MSH2*) and mutation types (point mutations versus large rearrangements) were assessed using χ^2 tests. Continuous data were reported in mean values and compared between groups using Student's *t* test. Comparisons are reported as odds ratios (OR) with 95% confidence intervals (95% CI). A two-sided P value of <0.05 was considered statistically significant.

Multivariable logistic regression analysis was used to further assess the associations of clinical features with mutation status and type. The multivariable analysis of type of mutation controlled for the presence of a mutation in the *MLH1* or *MSH2* gene. To account for all Lynch syndrome–related cancers, it was determined *a priori* that the presence of endometrial cancer and any other Lynch syndrome–related cancer(s) would be included in the model irrespective of statistical significance in the univariate analysis.

Results

Univariate Analysis. Two hundred eighty-five (15%) patients had detectable deleterious mismatch repair mutations in the *MLH1* or *MSH2* genes (112 *MLH1* and 173 *MSH2*). The clinical characteristics of the identified mutation carriers appear in Table 1. Additional

demographic data pertaining to ancestry and country of origin have been previously reported (9). Ninety-eight percent of the samples and data provided were by individuals residing in the United States. Only 63% (179 of 285) of all mutation carriers fulfilled the Modified Amsterdam Criteria. Conversely, the mismatch repair mutation detection rate was 34% (179 of 539) among all subjects fulfilling the Modified Amsterdam Criteria in the entire study population. *MSH2* mutation carriers were more likely to have family histories fulfilling the Amsterdam criteria than *MLH1* carriers (OR, 1.7; 95% CI, 1.04-2.8). Mutation carriers reported a total of 936 family members with colorectal cancer or other Lynch syndrome–related cancers.

Colorectal Cancer. The prevalence of colorectal cancer was higher in *MLH1* than *MSH2* mutation carriers (79% versus 69%, $P = 0.08$), but the percentage of mutation carriers with two or more colorectal cancer diagnoses was similar (14% versus 12% for *MLH1* and *MSH2* carriers, respectively; Table 1). Among *MLH1* carriers, 55 of 112 (50%) reported colorectal cancer as their only cancer compared with 74 of 173 (43%) of *MSH2* carriers ($P = 0.3$). *MLH1* carriers were significantly more likely to be diagnosed with colorectal cancer at age less than 50 years compared with *MSH2* carriers ($P = 0.01$). The mean age of colorectal cancer diagnosis in the entire sample was 42.2 years for *MLH1* carriers compared with 44.8 years for *MSH2* carriers (Table 2) where male *MLH1* carriers had a significantly younger mean age at colorectal cancer diagnosis compared with *MSH2* carriers (38 versus 44 years; $P < 0.01$). This age differential was not appreciated among female carriers.

The mean number of colorectal cancer tumors per family was significantly higher in families with *MLH1* than *MSH2* mutations (1.8 versus 1.5, respectively; $P = 0.04$, Table 1). In addition, relatives (FDR and SDR) of *MLH1* carriers reported earlier age at colorectal cancer diagnosis compared with relatives of *MSH2* carriers (mean age at colorectal cancer diagnosis: 40.5 versus 44.6 years; $P = 0.01$; Table 2).

Endometrial Cancer. Endometrial cancer was reported in 41% of all mutation carriers (68 of 167) and its prevalence did not statistically differ among females with *MLH1* or *MSH2* mutations (36% versus 44%, respectively). Of all females undergoing predictive genetic testing, 9% (119 of 1,308) reported endometrial cancer as their only cancer diagnosis; 16% (19 of 119) of these women were found to be *MSH2* mutation carriers and 6% (7 of 119) were *MLH1* mutation carriers ($P = 0.2$). The mean age at the time of endometrial cancer diagnosis

Table 1. Clinical characteristics of mutation carriers: *MLH1* versus *MSH2*

	Gene mutation		*P*	OR (95% CI)*
	MLH1 (*n* = 112)	*MSH2* (*n* = 173)		
	Frequency (%)	Frequency (%)		
Gender			0.39	0.8 (0.5-1.3)
Male	50 (45)	68 (39)		
Female	62 (55)	105 (61)		
Modified Amsterdam	62 (55)	117 (68)	0.03	1.7 (1.04-2.8)
Personal cancer history				
CRC				
Yes	88 (79)	119 (69)	0.08	0.6 (0.3-1.0)
No	24 (21)	54 (31)		
None	25 (21)	54 (31)	0.25	—
One	72 (64)	99 (57)		
Two or more	16 (14)	20 (12)		
CRC diagnosed at age <50 y	72 (83)	80 (67)	0.01	0.4 (0.2-0.8)
Endometrial cancer (among females only)	22/62 (36)	46/105 (44)	0.33	1.4 (0.7-2.7)
Other HNPCC†	10 (9)	41 (24)	0.001	3.2 (1.5-6.6)
Multiple HNPCC	33 (29)	51 (29)	1.00	1.0 (0.6-1.7)
Family cancer history				
Mean number of tumors				
FDR				
CRC	1.33	1.19	0.30	
Endometrial cancer	0.24	0.30	0.43	
Other HNPCC	0.14	0.39	<0.001	
SDR				
CRC	0.99	0.67	0.01	
Endometrial cancer	0.23	0.13	0.12	
Other HNPCC	0.13	0.16	0.53	
FDR + SDR‡				
CRC	1.82	1.53	0.04	
Endometrial cancer	0.36	0.36	0.99	
Other HNPCC	0.20	0.46	<0.001	

Abbreviation: CRC, colorectal cancer.
*Reference group: *MLH1* carriers (OR, 1.0).
†HNPCC includes cancers in the colorectum, kidney, ureter, bladder, brain, biliary tract, stomach, small intestine, ovary, pancreas, and sebaceous neoplasms.
‡A weighted sum of cancer diagnoses among FDR + SDR is used to incorporate the effect of genetic distance. Cancer diagnoses in SDR were weighted half that in FDR (12).

Table 2. Age of diagnosis among mutation carriers

	Gene mutation		P
	MLH1	*MSH2*	
Mean youngest age of diagnosis (±SE), y			
Carrier			
CRC	42.2 (±0.9)	44.8 (±0.7)	0.03
Endometrial cancer	45.8 (±0.5)	45.5 (±0.4)	0.69
Other HNPCC* cancer	49.1 (±3.2)	49.6 (±2.0)	0.92
Relatives			
FDR			
CRC	41.1 (±1.3)	44.8 (±1.1)	0.03
Endometrial cancer	48.0 (±1.2)	46.2 (±1.5)	0.38
Other HNPCC	49.9 (±3.1)	50.0 (±2.0)	1.00
SDR			
CRC	43.6 (±1.7)	52.2 (±1.8)	<0.01
Endometrial cancer	48.3 (±2.2)	48.4 (±3.0)	1.00
Other HNPCC	48.8 (±3.7)	50.9 (±3.9)	0.72
FDR + SDR†			
CRC	40.5 (±1.2)	44.6 (±1.1)	0.01
Endometrial cancer	46.8 (±1.1)	45.8 (±1.3)	0.55
Other HNPCC	49.2 (±2.4)	50.2 (±1.9)	0.77

NOTE: Carriers of *MLH1* and *MSH2* gene mutations.
*HNPCC includes cancers in the colorectum, kidney, ureter, bladder, brain, biliary tract, stomach, small intestine, ovary, pancreas, and sebaceous neoplasms.
†A weighted sum of cancer diagnoses among FDR + SDR is used to incorporate the effect of genetic distance. Cancer diagnoses in SDR were weighted half that in FDR (12).

was also similar for both *MLH1* and *MSH2* carriers, 45.8 versus 45.5 years, respectively (Table 2). There was no difference in the mean number of endometrial cancers or the mean age of cancer diagnosis among family members.

Other Lynch syndrome–associated cancers. Eighteen percent (51 of 285) of mutation carriers reported other Lynch syndrome–associated cancers. The prevalence of such cancers was significantly higher among *MSH2* carriers compared with *MLH1* carriers: 24% versus 9%, respectively (OR, 3.2; 95% CI: 1.5-6.6; Tables 1 and 3).

The proportion of carriers with history of sebaceous skin tumors differed significantly between *MLH1* and *MSH2* mutation carriers, 2% versus 8% (P = 0.05). Eighty-seven percent (13 of 15) of mutation carriers with sebaceous skin tumors had a *MSH2* mutation. Most of the urinary tract cancers (11 of 15) and ovarian cancers (11 of 14) were found in subjects with *MSH2* mutations. All other cancer types associated with Lynch syndrome were uncommon (Table 3). The mean number of other Lynch syndrome–related tumors per family (FDR + SDR) was lower in *MLH1* compared with *MSH2* carriers (P < 0.001; Table 1).

Type of Deleterious Mutation: Large Genomic Rearrangements versus Point Mutations. A total of 42 patients had large genomic rearrangements among 1,016 probands who underwent both full gene sequencing and large rearrangement analysis, with 7 in the *MLH1* and 35 in the *MSH2* genes. The age at diagnosis of colorectal cancer was the most striking difference in phenotype among carriers (48.3 years in those with large rearrangements versus 43.0 years for those with point mutations,

Table 3. Lynch syndrome-associated cancers among mutation carriers

Type of tumor	*MLH1* (n = 112)	*MSH2* (n = 173)	P
	Frequency (%)	Frequency (%)	
Colorectal	88 (79)	119 (69)	0.08
Endometrial	22 (20)	46 (27)	0.33
Other*	10 (9)	41 (24)	0.001
Urinary tract tumors	4	11	
Renal/kidney	1	6	
Ureter	2	3	
Bladder	1	4	
Brain	0	2	
Biliary	1	0	
Stomach	0	3	
Small intestine/duodenum	0	3	
Ovary	3	11	
Sebaceous†	2	13	
Pancreas	0	0	

*Four *MSH2* mutation carriers reported more than one Lynch syndrome–related cancer other than colorectal or endometrial cancer.
†P = 0.05.

Table 4. Clinical characteristics of carriers by type of mutation: point mutations versus large rearrangements

	Type of mutation		P	OR (95% CI)*
	Point mutations (n = 243)	Large rearrangements (n = 42)		
	Frequency (column %)	Frequency (column %)		
Gender			0.24	1.5 (0.8-2.9)
Male	97 (40)	21 (50)		
Female	146 (60)	21 (50)		
Modified Amsterdam	148 (61)	31 (76)	0.08	2.0 (0.9-4.2)
Personal cancer history				
CRC				
Yes	173 (71)	34 (81)	0.26	1.7 (0.8-3.9)
No	70 (29)	8 (19)		
None	71 (29)	8 (19)	0.14	—
One	139 (57)	31 (74)		
Two or more	33 (14)	3 (7)		
CRC diagnosed at age <50 y	135 (78)	17 (50)	0.001	0.3 (0.1-0.6)
Endometrial cancer (among females only)	58/146 (40)	10/21 (48)	0.49	1.4 (0.6-3.5)
Other HNPCC[†]	40 (16)	11 (26)	0.13	1.8 (0.8-3.9)
Multiple HNPCC	70 (29)	14 (33)	0.58	1.2 (0.6-2.5)
Family cancer history				
Mean number of tumors:				
FDR				
CRC	1.23	1.33	0.47	
Endometrial cancer	0.28	0.26	0.86	
Other HNPCC	0.26	0.45	0.11	
SDR				
CRC	0.79	0.81	0.92	
Endometrial cancer	0.15	0.24	0.47	
Other HNPCC	0.13	0.21	0.20	
FDR + SDR[‡]				
CRC	1.63	1.74	0.45	
Endometrial cancer	0.35	0.38	0.79	
Other HNPCC	0.33	0.56	0.07	

NOTE: Carriers of *MLH1* and *MSH2* gene mutations.
*Reference group: *MLH1* carriers (OR, 1.0).
†HNPCC includes cancers in the colorectum, kidney, ureter, bladder, brain, biliary tract, stomach, small intestine, ovary, pancreas, and sebaceous neoplasms.
‡A weighted sum of cancer diagnoses among FDR + SDR is used to incorporate the effect of genetic distance. Cancer diagnoses in SDR were weighted half that in FDR (12).

P = 0.001; Table 5). Individuals with point mutations were more likely to have colorectal cancer diagnosed at age less than 50 years (Table 4).

The mean number of colorectal and endometrial tumors among FDRs and SDRs was similar between the two mutation types. However, family members of mutation carriers with large rearrangement mutations had a slightly higher mean number of other Lynch syndrome–related cancers (not statistically significant; Table 4). Overall, the mean youngest age of any Lynch syndrome cancer (including colorectal cancer) among FDRs and SDRs did not differ between carriers with point mutations or large rearrangements (Table 5).

Multivariable Analysis. Multivariable logistic regression analysis identified three clinical features predictive of *MLH1* versus *MSH2* mutation: (*a*) *MLH1* mutation carriers had a slightly higher likelihood of having more colorectal cancer diagnosed among family members than *MSH2* carriers (OR, 1.3; 95% CI, 1.0-1.7); (*b*) *MSH2* mutation carriers were more likely to have a personal history of nonendometrial extracolonic cancers compared with *MLH1* mutation carriers (OR, 3.2; 95% CI, 1.4-7.3); and (*c*) *MSH2* mutation carriers were more likely to have an increased number of other cancers diagnosed among relatives than *MLH1* mutation carriers (OR, 2.1; 95% CI, 1.2-3.7). Personal history of endometrial cancer and colorectal cancer diagnosed at a young age were not predictive of which mismatch repair gene was mutated (Table 6). No significant interactions were found between the independent predictors and age or gender. There were also no clinical variables predictive of the type of mutation (Table 6).

Discussion

Our study provides data on genotype-phenotype relationships associated with germ-line mutations in the mismatch repair genes *MLH1* and *MSH2* from a large, diverse population of mutation carriers from the United States. The results shed light on important similarities and differences between other studies of mismatch repair gene mutation carriers, which have been conducted primarily in European registry–based populations (Table 7). Because of the large number of mutation carriers, we were also able to evaluate whether the type of mutation affected phenotypic expression and found results contrary to our expectations.

Compared with *MSH2* carriers, *MLH1* mutation carriers were more likely to have colorectal cancer as

Table 5. Age of diagnosis among carriers with point mutations versus large rearrangements

	Type of mutation		P
	Point mutations	Large rearrangements	
Mean youngest age of diagnosis (±SE), y			
Carrier			
CRC	43.0	48.3	0.001
Endometrial cancer	45.7	44.7	0.12
Other HNPCC* cancer	49.7 (±2.1)	48.7 (±2.8)	0.82
Relatives			
FDR			
CRC	42.8 (±0.9)	45.9 (±2.3)	0.14
Endometrial cancer	46.8 (±1.2)	47.3 (±1.6)	0.85
Other HNPCC	50.9 (±1.9)	46.0 (±3.4)	0.24
SDR			
CRC	47.6 (±1.4)	52.7 (±3.6)	0.17
Endometrial cancer	48.7 (±2.1)	46.0 (±5.3)	0.64
Other HNPCC	48.5 (±2.8)	57.2 (±8.7)	0.23
FDR + SDR†			
CRC	42.8 (±0.9)	44.1 (±2.1)	0.54
Endometrial cancer	46.1 (±1.0)	46.3 (±1.8)	0.94
Other HNPCC	50.0 (±1.6)	49.5 (±3.6)	0.89

NOTE: Carriers of *MLH1* and *MSH2* gene mutations.

*HNPCC includes cancers in the colorectum, kidney, ureter, bladder, brain, biliary tract, stomach, small intestine, ovary, pancreas, and sebaceous neoplasms.

†A weighted sum of cancer diagnoses among FDR + SDR is used to incorporate the effect of genetic distance. Cancer diagnoses in SDR were weighted half that in FDR (12).

the only type of cancer diagnosed, an increased number of colorectal cancer among relatives, and a younger age of cancer diagnosis. The younger age at diagnosis of colorectal cancer was more notable among *MLH1* mutation–positive male carriers than in females. Contrary to prior findings in other studies, we found no appreciable difference in endometrial cancer prevalence or age of diagnosis between *MLH1* and *MSH2* mutation carriers. Similar to our previously reported experience (13) and studies from European registry–based populations, other Lynch syndrome–associated cancers were more prevalent among both *MSH2* mutation carriers and their relatives. Overall, we found that a personal history or an increased number of extracolonic cancers other than endometrial among family members were clinical features associated with the presence of a *MSH2* gene mutation.

We evaluated the effect of the type of gene alteration on phenotype, expecting that large rearrangements and deletions would lead to a more severe phenotype with more multiple cancers and earlier ages of onset (14). The results unexpectedly showed that individuals who had a rearrangement had older ages at cancer diagnosis but age was not found to be a significant independent predictor of mutation type. Overall, there were few differences between the cancer histories of individuals with point mutations versus large rearrangements, which suggest that the specific mismatch repair gene involved is more important than the type of mutation in determining phenotype.

Our results illustrate some important similarities and differences among reports of genotype-phenotype associations in Lynch syndrome. Reviewing the data from relatively large cohorts of Dutch, Finnish, French, and German *MSH2* and *MLH1* mutation carriers (7, 15-18), most studies have reported that extracolonic nonendometrial cancers are more prevalent in *MSH2* carriers than in *MLH1*. Although the small numbers of these cancers in this study limited statistical power, ovarian and upper urinary tract cancers were slightly more prevalent among *MSH2* mutation carriers, corroborating previous reports by Vasen et al. (7). Our finding that sebaceous skin tumors were significantly associated with *MSH2* mutations also supports data from Mangold et al. (19)

Table 6. Multivariable analysis for factors predicting gene mutation status and mutation type

Characteristics	Odds ratio (95% CI)	
	*MLH1** vs *MSH2*	Point mutations* versus large rearrangements
CRC age of diagnosis less than 50 y	0.6 (0.3-1.3)	0.7 (0.3-1.5)
Endometrial cancer in carriers	1.7 (0.8-3.4)	0.8 (0.4-1.9)
Extracolonic cancers† in carriers	3.2 (1.4-7.4)	1.3 (0.6-2.9)
Increased number of CRC tumors in FDR+SDR‡	0.8 (0.6-0.9)	1.1 (0.9-1.6)
Increased number of extracolonic tumors in FDR+SDR	2.1 (1.2-3.6)	1.4 (0.9-2.2)

*Reference group.

†Extracolonic cancers include kidney, ureter, bladder, brain, biliary tract, stomach, small intestine, ovary, pancreas, and sebaceous neoplasms.

‡A weighted sum of cancer diagnoses among FDR + SDR is used to incorporate the effect of genetic distance. Cancer diagnoses in SDR were weighted half that in FDR (12).

Table 7. Summary of previous studies comparing genotype and phenotype in Lynch syndrome

Study	Population	No. MMR mutations	No. subjects (families/ mutation carriers)	Summary findings of phenotypic comparison
Vasen et al. (7)	Dutch	34 *MLH1* 40 *MSH2*	79 families 1,842 related carriers	CRC: higher lifetime risk in *MSH2* (NS) Endometrial cancer: higher lifetime risk in *MSH2* (NS) Other cancers: brain, stomach, ovarian, urinary tract* more prevalent in *MSH2*
Peltomaki et al. (16)	Finnish	51 *MLH1* 4 *MSH2*	55 families 295 related carriers	CRC: more prevalent in *MLH1* Endometrial cancer: more prevalent in *MSH2* (NS) Other cancers: more prevalent in *MSH2* (NS)
Parc et al. (17)	French	65 *MLH1* 79 *MSH2*	163 index cases 348 related carriers	CRC: no difference Endometrial cancer: higher risk in *MSH2* (NS) Other cancers: no difference
Goecke et al. (18)	German	124 *MLH1* 157 *MSH2*	281 families 988 related carriers	CRC: more prevalent in *MLH1**; younger age of diagnosis in *MLH1* males Endometrial cancer: no difference Other cancers: sebaceous skin tumors,* prostate, bladder more prevalent in *MSH2*
Kastrinos et al.	United States	112 *MLH1* 173 *MSH2*	285 unrelated carriers 936 affected family members	CRC: more prevalent in *MLH1* (NS); younger age of diagnosis in *MLH1* males Endometrial cancer: no difference Other cancers: urinary tract, ovarian, sebaceous skin tumors* more prevalent in *MSH2*

Abbreviations: MMR, mismatch repair; NS, statistically nonsignificant.
*$P < 0.05$.

regarding the increased prevalence of *MSH2* mutations among individuals with the Muir-Torre variant of Lynch syndrome.

The endometrial cancer prevalence results have shown more variability. A majority of prior studies, some of which did not distinguish endometrial cancer from other extracolonic cancers, report an increased prevalence of endometrial cancer in *MSH2* mutation carriers. Most recently, this association has not been supported (18). Therefore, in combination with our results, it is becoming apparent that endometrial cancer rates and ages of onset do not differ between *MLH1* and *MSH2* mutation carriers.

Finally, the earlier age of onset of colon cancer that is more prevalent in male *MLH1* mutation carriers is an interesting finding. The French study suggested this gender difference although it was not statistically significant (17). Only recently have Goecke and the German HNPCC Consortium reported a similar association in their analysis of 988 related subjects in 281 mutation-positive families (18). The biological basis for this gender difference is not clear and warrants further study.

An important strength of our study is that it represents a diverse population of unrelated individuals who were not selected from family registries. It represents the largest number of unrelated mutation carriers studied to date. The majority of subjects did not meet the strict clinical criteria that have previously been used to select persons suitable for mutation analysis. Previous studies have specifically analyzed mismatch repair gene mutations in selected kindreds with familial clustering of colorectal cancer. Therefore, the ascertainment of the families not only leads to an overestimation of colorectal cancer risk but also hampers the ability to define a reliable phenotypic association to particular germ-line mismatch repair gene mutations. Ethnic homogeneity among certain European cohorts may highlight a founder mutation effect leading to an overrepresentation of certain extracolonic cancers. In addition, environmental factors may contribute to the differences seen in the prevalence of extracolonic cancers. For example, the higher incidence of gastric cancer in the general populations of Finland compared with the Dutch population may suggest an environmental contribution to the differences seen in the gastric cancer incidence between the Dutch and the Finnish families (6, 7).

Nevertheless, potential limitations of the current analysis must be acknowledged. This study is not truly ''population-based.'' The least biased of all possible samples would involve population screening of patients with sporadic cancer, which represents an enormous, expensive undertaking that may provide very low-yield results with respect to the number of mismatch repair gene mutations detected. In this study, subjects were selected to undergo testing based on some personal or family clinical history that triggered a heightened suspicion for Lynch syndrome. This population is typical of one that is referred to cancer genetics clinics for consultation, counseling, and predictive genetic testing, and best represents a mutation frequency more reflective of the general population at risk.

In addition, the cross-sectional prevalence study design does not afford us the ability to determine whether a prognostic advantage exists between the two mismatch repair gene mutation carriers. Analyzing data provided by surviving mutation carriers at one distinct point in time may influence estimates of the overall prevalence of disease and the comparisons made between the two mutation carrier groups. In turn, the conclusions drawn from this cross-sectional prevalence study may be vulnerable to inferential reasoning. Prospective, population-based studies of a large number of mutation carriers would be necessary to completely resolve such a survival bias and such studies are difficult to undertake due to the rarity of Lynch syndrome.

A third limitation to this study is the inability to confirm the cancer diagnoses reported in probands and family members. Although previous evidence shows that family history reported by probands in FDRs is reliable, the same cannot be said for SDRs. Reporting errors are possible in this study. However, because health care professionals were responsible for the clinical data provided, it is less likely that erroneous diagnoses were documented. Despite these shortcomings, the strong predictive effects as reported in our previous work using this study group (9) that have recently been validated in a population-based sample (20) are indicative that the information is likely to be reliable.

In conclusion, the phenotypic expression of disease in individuals and within families with Lynch syndrome is influenced by whether the genetic mutation is present in *MLH1* or *MSH2*. In the United States, a preponderance of colorectal cancer is more often associated with a mutation in the *MLH1* gene with an increased number of colorectal tumors seen among relatives of *MLH1* mutation carriers. Extracolonic Lynch syndrome–associated tumors, other than endometrial cancer, predominate in *MSH2* carriers with a higher tumor burden noted among family members. Testing for a specific mismatch repair gene based on family history features may provide an efficient approach to predictive genetic testing. The clinical phenotypes described in this study could be used to tailor recommendations regarding cancer screening, surveillance, and prevention for patients and families with Lynch syndrome.

Disclosure of Potential Conflicts of Interest

Dr. Syngal reports having received lecture honoraria from Myriad Genetics Laboratories, Inc. No other conflicts of interest were disclosed.

Acknowledgments

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Myriad Genetics, Inc., for sharing the data for analysis.

References

1. Lynch HT, de la Chapelle A. Hereditary colorectal cancer. N Engl J Med 2003;348:919–32.
2. Grady WM. Genetic testing for high-risk colon cancer patients. Gastroenterology 2003;124:1574–94.
3. Chung DC, Rustgi AK. The hereditary nonpolyposis colorectal cancer syndrome: genetic and clinical implications. Ann Intern Med 2003;138:560–70.
4. Aarnio M, Sankila R, Pukkala E, et al. Cancer risk in mutation carriers of DNA-mismatch-repair genes. Int J Cancer 1999;81:214–8.
5. Watson P, Lynch HT. Extracolonic cancer in hereditary nonpolyposis colorectal cancer. Cancer 1993;71:677–85.
6. Aarnio M, Mecklin JP, Aaltonen LA, Nystrom-Lahti M, Jarvinen HJ. Life-time risk of different cancers in hereditary non-polyposis colorectal cancer (HNPCC) syndrome. Int J Cancer 1995; 64:430–3.
7. Vasen HFA, Stormorken A, Menko FH, et al. *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; 19:4074–80.
8. Lynch HT, Boland CR, Gong G, et al. Phenotypic and genotypic heterogeneity in the Lynch syndrome: diagnostic, surveillance and management implications. Eur J Hum Genet 2006;14:390–402.
9. Balmaña J, Stockwell DH, Steyerberg EW, et al. Prediction of *MLH1* and *MSH2* mutations in Lynch Syndrome. JAMA 2006; 296:1469–78.
10. Barnetson RA, Tenesa A, Farrington SM, et al. Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. N Engl J Med 2006;354:2751–63.
11. Chen S, Wang W, Lee S, et al. Prediction of germline mutations and cancer risk in the Lynch Syndrome. JAMA 2006;296:1469–78.
12. Steyerberg EW, Balmaña J, Stockwell DH, Syngal S. Data reduction for prediction: a case study on robust coding of age and family history for the risk of having a genetic mutation. Stat Med 2007;26: 5545–56.
13. Pandipalliam B, Garber J, Kolodner RD, Syngal S. Clinical presentation correlates with the type of mismatch repair gene involved in hereditary nonpolyposis colon cancer. Gastroenterology 2004;126: 936–7.
14. Baudhuin LM, Ferber MJ, Winters JL, et al. Characterization of h*MLH1* and h*MLH2* gene dosage alterations in Lynch Syndrome patients. Gastroenterology 2005;129:846–54.
15. Vasen HF, Wijnen JT, Menko FH, et al. Cancer risk in families with hereditary nonpolyposis colorectal cancer diagnosed by mutation analysis. Gastroenterology 1996;111:1020–7.
16. Peltomaki P, Gao X, Mecklin JP. Genotype and phenotype in hereditary nonpolyposis colon cancer: a study of families with different vs. shared predisposing mutations. Fam Cancer 2001;1: 9–15.
17. Parc Y, Boisson C, Thomas G, Olschwang S. Cancer Risk in 348 French *MSH2* or *MLH1* gene carriers. J Med Genet 2003;40: 208–13.
18. Goecke T, Schulmann K, Engel C, et al. Genotype-phenotype comparison of German *MLH1* and *MSH2* mutation carriers clinically affected with Lynch Syndrome: a report by the German HNPCC Consortium. J Clin Oncol 2006;24:4285–92.
19. Mangold E, Pagenstecher C, Leister M, et al. A genotype-phenotype correlation in HNPCC: strong predominance of *MSH2* mutations in 41 patients with Muir-Torre syndrome. J Med Genet 2004;41: 567–72.
20. Balaguer F, Balmaña J, Castellví-Bel S, et al. Validation and extension of the PREMM$_{1,2}$ model in a population-based cohort of colorectal cancer patients. Gastroenterology 2008;134:39–46.

7.4. Appendix 4. “Mismatch repair gene analysis in Catalonian families with colorectal cancer”.

M.Palicio, J.Balmaña, S.González, I.Blanco, E.Marcuello, MA. Peinado, G.Julià, JR. Germà, JJ. López López, J. Brunet, G. Capellà.

J Med Genet 2002; 39:e29

ELECTRONIC LETTER

Mismatch repair gene analysis in Catalonian families with colorectal cancer

M Palicio, J Balmaña, S González, I Blanco, E Marcuello, M A Peinado, G Julià, J R Germà, J J López López, J Brunet, G Capellà

J Med Genet 2002;**39**:e29 (http://www.jmedgenet.com/cgi/content/full/39/6/e29)

Colorectal cancer (CRC) is one of the most common malignant neoplasms in western countries.[1] Hereditary non-polyposis colorectal cancer (HNPCC) is thought to represent the most common form of familial colorectal carcinoma and may account for approximately 1-6% of all colorectal cancers.[2] The clinical phenotype of HNPCC is variable, and there is no distinctive clinical hallmark, such as the presence of more than 100 adenomatous polyps in familial adenomatous polyposis (FAP). Therefore, stringent diagnostic criteria, essentially based on personal and family cancer history, have been adopted for the purpose of identifying HNPCC families. The Amsterdam Criteria for HNPCC diagnosis include: (1) the presence of three or more patients affected with CRC or HNPCC related tumours (mainly endometrial cancer), two of whom must be first degree relatives (sib, parent, or offspring) to the other one; (2) vertical transmission of CRC in two successive generations, indicative of autosomal dominant inheritance; (3) early onset (<50 years) of at least one CRC or HNPCC related tumour in the family; and (4) exclusion of FAP.[3 4] However, Amsterdam criteria can be too restrictive when applied to small kindreds, abundant in western countries.[5] In addition, those cases likely to represent the first mutation in a family, usually single affected members diagnosed before the age of 45, may be missed. Finally, difficulty in eliciting a complete and accurate cancer family history may preclude widespread use of these clinical criteria.[6 7]

hMSH2 and *hMLH1* are considered to be the two major genes responsible for HNPCC.[8] Initially, germline mutations in these genes were detected in up to 90% of HNPCC families suggesting that the Amsterdam criteria were useful for selecting kindreds which were candidates for genetic analyses. Nowadays, there is general agreement that approximately 40-50% of families meeting the Amsterdam criteria have detectable *hMSH2* or *hMLH1* mutations.[2] Furthermore, some recent studies suggest that the mutation detection rate may be even lower (25%).[9 10] On the other hand, a significant number of germline mutations in the mismatch repair (MMR) genes have been detected in kindreds with clustering of CRC not meeting Amsterdam criteria.[11] The prevalence of these mutations in these families is usually lower than that reported for families meeting these criteria. Therefore, both the sensitivity and specificity of the Amsterdam criteria for selecting those patients who are candidates for MMR genetic analysis have been challenged.

The aim of this study was to evaluate the prevalence of germline *hMSH2* and *hMLH1* mutations in 11 families meeting the Amsterdam criteria and 21 HNPCC-like Spanish families (11 of them lacking one AC), and to characterise the mutation spectrum in north eastern Spain. We only detected mutations in the *hMLH1* gene, and the majority of mutations were in families lacking one Amsterdam criterion.

MATERIAL AND METHODS

Patients and families

A total of 60 kindreds with clustering of colorectal cancer were recruited from the Genetic Counselling Unit in Hospital Sant Pau, Barcelona, between September 1996 and November 1998. Most families were referred from other Catalonian hospitals. Genetic analysis was indicated when at least one Amsterdam criterion was present. Thirty-two families were included in a study to analyse the prevalence of *hMSH2* and *hMLH1* gene mutations in HNPCC and HNPCC-like families. Eleven of these families fulfilled the Amsterdam criteria[4] and the remaining 21 kindreds were defined as HNPCC-like: 11 lacked one Amsterdam criterion and the remaining 10 families lacked two Amsterdam criteria. Pedigrees were constructed up to three generations and data from affected subjects (type, number, and localisation of tumours, age at diagnosis, and pathological features) were collected if available. Data confirmation was obtained after reviewing medical or pathological records or death certificates. All patients eligible for genetic analysis received genetic counselling before and after the completion of the analysis, and informed consent was obtained before the study. The ethics review committee of Sant Pau Hospital approved this study.

Detection of germline mutations in *hMSH2* and *hMLH1* genes

Peripheral blood cells were stored at –20°C until DNA or RNA extraction. High molecular weight DNA was extracted following standard procedures. RNA extraction was performed using the single step method.[12] DNA from at least two healthy controls was analysed as negative controls in each experiment.

hMHS2 analyses

cDNA analyses

In the first 22 cases, mutation detection was performed by PCR/SSCP analyses of cDNA sequences. cDNA was obtained after RT using standard protocols.[13] The whole coding sequence of the gene was amplified in nine overlapping fragments using primers based upon the published cDNA sequence.[14] Primers used and PCR conditions are available from the authors upon request. SSCP analyses were performed in an ALF-Express apparatus (AmershamPharmacia-biotech) with at least two running conditions: (1) 12% PAGE at 30°C and 30 W for two hours, and (2) 6% PAGE in 10% glycerol at 30°C and 7 W for six hours. For exon 5 deletion analyses,[15] fragments 3 and 4 were amplified in a single PCR reaction (amplicon size 635 bp) and the presence of a second fragment of smaller size (485 bp) was searched for.

Abbreviations: CRC, colorectal cancer; HNPCC, hereditary non-polyposis colorectal cancer; FAP, familial adenomatous polyposis; MMR, mismatch repair; AC, Amsterdam criteria; MSI, microsatellite instability

DNA analysis

During the period of study of *hMSH2*, the material was changed from RNA to DNA in order to be able to analyse samples referred from other hospitals. Therefore, the whole coding exon sequences with corresponding intron boundaries of the remaining 10 samples were amplified using intronic primers (sequence available upon request to authors). SSCP analyses were run under the same two conditions described above. In this case, unlabelled primers and gel silver staining were used. Whenever an abnormal SSCP pattern was observed, direct sequencing of the amplified fragment was performed using the AmpliCycle Sequencing Kit (Perkin Elmer, Branchburg, NJ) following the manufacturer's instructions. Sequence analyses were performed with either a PE ABIPrism 670 or ALFExpress (AmershamPharmaciaBiotech) sequencer. More than 60% of the fragments amplified showing a normal SSCP pattern were randomly selected for sequencing in order to reinforce the reliability of our negative results.

hMLH1 analyses

hMLH1 gene mutation detection was performed on DNA because of the presence of alternate splicing that could make interpretation of results difficult.[16] All exons of the *hMLH1* gene with their corresponding exon-intron boundaries were analysed by PCR/SSCP following essentially the same protocol as described for *hMSH2*. Primers used and PCR conditions are available from the authors upon request. Whenever an abnormal SSCP pattern was observed, direct sequencing of the amplified fragment was performed as described above. Owing to the detection of a significant number of mutations in the *hMLH1* gene, no sequencing was performed when normal SSCP patterns were observed.

Microsatellite (MSI) analysis

Attempts were made to collect paraffin blocks of affected members from all families studied. Blocks were available from 22 of the 32 families analysed (five of 11 HNPCC families and 16 of 21 HNPCC-like families). Enrichment for tumour cells was performed in all cases. Paired results from tumour and corresponding normal mucosa were always assessed. The strategy for MSI diagnosis has been previously described.[17] Briefly, as a first step, two markers (one containing mono-runs of As BAT26 and one CA repeat D12S95) were assessed. If both markers were stable (no additional bands seen), the tumour was classified MSI(−). If two markers were unstable, the tumour was classified MSI(+) (sensitivity 97%, specificity 100%). Whenever only one of the two markers was unstable (20% of samples), an additional four markers were studied. Tumours were classified as MSI(+) when at least two of the six markers were unstable. Tumours exclusively displaying instability at BAT26 were considered highly suggestive of displaying MSI. Inconclusive results were reported whenever amplification was achieved in less than four markers.

RESULTS

Clinical and pathological features

As expected, age at diagnosis of CRC was lower in HNPCC than in HNPCC-like families. Right sided tumours represented 31% of all CRC in HNPCC families and only 18% of HNPCC-like ones, but the difference was not significant (table 1). Colorectal tumours were detected at similar stages (Dukes A and B) in both groups. No differences were observed regarding the incidence of synchronous and metachronous CRC between the two groups (data not shown).

Table 1 Clinical features of HNPCC and HNPCC-like families

	HNPCC (n=11)	HNPCC-like (n=21)	p
No of CRC	38	54	
Age at diagnosis	48 (22–80)	56 (24–89)	0.089
Localisation			
Right	12 (31.6%)	10 (18.5%)	NS
Left	15 (39.5%)	19 (35.2%)	
Rectum	6 (15.8%)	12 (22.2%)	
Unknown	5	13	
Stage (Dukes)			
A/B	12 (31.6%)	21 (38.9%)	NS
C/D	16 (42.1%)	11 (20.4%)	
Unknown	10	22	
Differentiation*			
W/M	27 (75.1%)	32 (59.3%)	NS
P	2 (5.3%)	4 (7.4%)	
Unknown	5	15	
Synchronous CRC	2 (5%)	3 (5.5%)	NS

*W: well differentiated; M: moderately differentiated; P: poorly differentiated.

Germline analyses

Three of the 11 (27%) HNPCC families harboured a MMR gene mutation, exclusively in the *hMLH1* gene (table 2). One of these three *hMLH1* mutations is a previously reported deletion AAG at 1846 (exon 16, codon 616).[18 19] The remaining two mutations are novel and of unknown pathogenicity. Family J/8 harbours a mutation located at position +1 of intron 8 resulting in a disrupted splicing donor site according to Spliceview software (htpp://125.itba.mi.cnr.it/ genebin/ wwwspliceview). The third mutation is a Leu to His substitution at codon 622 in exon 16 that was detected in two affected members of family J/35. Unfortunately, we have not been able to collect tumour tissue from affected members of the family. We are currently trying to contact additional relatives of the latter families to complete segregation analysis. It is of note that no exon 5 deletion, probably the most frequently reported mutation, was detected despite using a sensitive method to detect this mutation.

A similar proportion of HNPCC-like families (four of 21, 19%) harboured *hMLH1* gene mutations. Interestingly, all these mutations were identified in families that lacked only one Amsterdam criterion resulting in a four of 11 (36%) mutation yield in this subgroup. Two of the four *hMLH1* mutations, both single base substitutions, have been previously described as deleterious (table 2). Accordingly, analysed tumours were MSI(+). A novel mutation, one 4 bp insertion generating a stop codon at nt 1664, is most likely to be deleterious and is associated with a MSI(+) tumour. Finally, a 5 bp deletion located intronically, which apparently does not affect splicing, was detected in family J/39. The fact that the tumour sample analysed was classified as MSI(−) further raises doubts about its pathogenicity.

Since mutations were similarly distributed among HNPCC and HNPCC-like families, the usefulness of individual criteria for predicting the presence of a germline mutations was analysed. No significant differences were observed between (1) vertical transmission, (2) the presence of three or more affected members, (3) age at diagnosis under 50 in at least one affected subject, and (4) the presence of detectable mutations (data not shown). When families were divided according to the presence or absence of detected mutations, no differences were observed regarding clinical and pathological parameters evaluated, including age of onset and the presence of endometrial cancer (data not shown).

Variant alleles

A total of six polymorphisms (three in the *hMSH2* and three in the *hMLH1* genes) were detected (table 3), all of them previously reported. It is of note that an *hMSH2* intron 1 polymorphism was detected in 12 of the 33 (36%) families. Three other polymorphisms at intron 12 of *hMSH2* and introns 8 and 13 of *hMLH1* were detected in three families each.

Table 2 Mutations in the *hMSH2* and *hMLH1* genes in HNPCC and HNPCC-like families

Family	Gene	Location	Nature	Consequence	Pathogenicity	References
HNPCC						
J/8	*hMLH1*	Intron 8	G>A; nt 677+1	Affects splice donor site	Unknown	Novel
J/42	*hMLH1*	Exon 16, codon 616	Del AAG; nt 1846	Frameshift	Deleterious	Hamilton *et al*[18] Miyaki *et al*[19]
J/35	*hMLH1*	Exon 16, codon 622	T>A; nt 1865	Single AA substitution Leu>Hys	Unknown	Novel
HNPCC-like						
J/39	*hMLH1*	Intron 1	del AGTAG; nt 207; –44	Does not affect splicing	Unknown	Novel
J/9	*hMLH1*	Exon 2, codon 67	G>A; nt 199;	Single AA substitution Gly>Arg	Deleterious	Tannergard *et al*[29]
J/38	*hMLH1*	Exon 14, codon 555	Ins AAGT; nt 1664	Stop codon	Probably deleterious	Novel
J/56	*hMLH1*	Exon 19, codon 714	G>A; nt 2141	Stop codon	Deleterious	Froggatt *et al*[30] Hutter *et al*[31]

Table 3 Intragenic variant alleles in the *hMSH2* and *hMLH1* genes in HNPCC and HNPCC-like families

Exon	Location	Nature	Affected families (No)	References
hMLH1 gene				
Exon 2	Codon 66	C>T nt198; Thr>Thr	1	Wehner *et al*[32]
Exon 8	Codon 219	A>G nt 655; Ile>Val	3	Liu *et al*[25]
Intron 13	3′ exon 13	G>A nt1558 +14	3	Tannergard *et al*[29]
hMSH2 gene				
Intron 1	3′ exon 1	C>G nt 211 +9	12	Bubb *et al*[33]
Exon 2	Codon 113	G>A nt; Lys>Lys	1	Liu *et al*[25]
Intron 12	5′ exon 13	T>C nt 2006 –6	3	Kolodner *et al*[34] Leach *et al*[15]

Concordance between MSI status and the presence of germline mutations

Tumours from patients belonging to 21 families were available for study (table 4). Six of the 21 (29%) tumours analysed were MSI(+), five of 16 belonging to HNPCC-like families and one of five HNPCC families, and one tumour was classified as highly suspicious of being MSI(+). As previously described, a strong correlation was observed between MSI status and detectable mutations. Mutations were detected in four of six MSI(+) tumours and only one substitution, of unknown pathogenicity, was detected in the 11 MSI(−) cases. If only MSI(+) or inconclusive cases (eight cases) had been analysed for germline mutation, the yield would have been 50% (four of eight).

Table 4 Correlation between MSI status and *hMLH1* gene mutations in available tumours

Family			
HNPCC	HNPCC-like	MSI status	*MLH1* gene
J/6		Negative	WT
J/40		Negative	WT
J/41		Negative	WT
J/42		Positive	del AAG; nt 1846
J/43		Negative	WT
	J/9	Positive	G>A; nt 199
	J/11	Negative	WT
	J/12-13	Negative	WT
	J/27	Highly suggestive	WT
	J/29	Positive	WT
	J/34	Inconclusive	WT
	J/36	Negative	WT
	J/38	Positive	Ins AAGT; nt 1664
	J/39	Negative	del AGTAG; nt 207
	J/39	Negative	WT
	J/44	Negative	WT
	J/46	Inconclusive	WT
	J/53	Negative	WT
	J/55	Negative	WT
	J/56	Positive	G>A; nt 2141
	J/57	Positive	WT

DISCUSSION

In the present study, we have shown that more than 50% of *hMLH1* putative germline mutations identified in our kindreds were detected in HNPCC-like families lacking one Amsterdam criterion. It has been reported that MMR gene mutations, although not restricted to HNPCC families, occurred at a higher frequency than in HNPCC-like families.[20] In our series, the mutation rate in families lacking one criterion is slightly higher than in HNPCC families, in line with observations by other authors.[10 21] The mutation rate in *hMSH2* and *hMLH1* genes in HNPCC families is usually around 40-60% (ranging from 22% to 86%). The observed frequency (27%) of mismatch repair gene mutations in our Mediterranean area is among the lowest reported. Several factors may account for the wide variations observed in distinct series. A first possibility is that significant differences may be present in different geographical areas.[5 22] We have previously shown that MSI(+) tumours represent a small proportion (7.5%) of Spanish CRC tumours[17] when compared to the USA (up to 15-20%). However, the same low frequency of MMR gene mutations was observed in German HNPCC families.[10] Secondly, differences in the methodology may also be relevant. However, most studies, including those obtaining the most discordant results, used SSCP as the initial screening technique.[11 23 24] While this technique may overlook some mutations, it is a well accepted method for point mutation screening. It is of note that a recent report has suggested that large deletions, as assessed by Southern blotting, may account for up to 10% of all *hMSH2* mutations.[25] However, since none of the previously mentioned reports used this methodology, all of them, including ours, may have underestimated the total number of MMR gene mutations in a similar manner. The incidence of MSI(+) tumours (31%) in our families is consistent with a good sensitivity of the techniques used. Large deletions in *hMSH2* or

hMLH1 and/or mutations in other genes (*hMSH6*) are likely to account for those MSI(+) tumours not showing detectable germline mutations. Finally, a consistent trend towards a lower proportion in MMR(+) cases in later publications may reflect the presence of some positive case bias in early reports probably associated with a larger number of affected subjects per family in the initially selected families. In this regard, recent results may offer a more realistic estimation of the impact of MMR gene analysis in the genetic counselling of these patients.

In our series, like others,[11] fulfilment of Amsterdam criteria has shown limited value in the identification of MMR gene mutation carriers. Our results favour the use of two Amsterdam criteria other than exclusion of FAP as a good cut off value for the selection of kindreds that are candidates for genetic analysis. If these clinical criteria are combined with MSI analysis, then the results of MMR analysis are even better, further supporting the combined use of clinical and molecular criteria to select patients for germline analysis.[26]

The spectrum of mutations shows that *hMLH1* mutations account for the majority of HNPCC germline alterations in this population. Since it was first described, an increasing role for *hMLH1* mutations in HNPCC has been observed.[2] In our series we only found *hMLH1* mutations, either deleterious or of unknown significance. No *hMSH2* gene mutations were detected in spite of intensive efforts aimed at minimising the rate of false negative results. Neither extensive use of sequencing nor the development of specific strategies for detection of exon 5 deletions has increased the total number of *hMSH2* mutations detected. Altogether our observations suggest that in the Spanish population *hMLH1* should be screened first for mutations and then *hMSH2* should be searched for alterations. Analysis of more families will most likely result in the occasional identification of *hMSH2* germline mutations in our setting.

Neither a founder effect nor hot spots nor recurrent mutations were observed, contrary to what occurs in Finland.[5] Heterogeneity of mutations has been high with a significant proportion (four of seven) of novel mutations. As usual, the significance of some novel alterations is equivocal especially when they are located in intronic regions or they are single base substitutions. In this situation, segregation analysis may be of help but the difficulty in obtaining biological material of other relatives has precluded it. MSI analysis may also be helpful in interpretation of results: the deletion at intron 1 is associated with a MSI(−) tumour disregarding a putative pathogenetic role. Regarding variant alleles, it is of note that *hMSH2* intron 1 polymorphism has been detected in 12 of the 33 (36%) families, further supporting the possibility that this variant allele confers an increased risk for developing familial CRC. Further studies will be needed to rule out this possibility. Finally, it is of note that, in our series, the prevalence of mutations in families meeting classical and revised Amsterdam criteria (that give extracolonic HNPPC tumours the same diagnostic rank as CRC) was identical, reinforcing the adequacy of modifying the criteria.

The total low number of MMR gene mutations detected strongly suggests that other genes[27] will be responsible for the increased CRC predisposition observed in those HNPCC families not harbouring MMR mutations.[2 5] Whether a family fulfils the clinical criteria or not, a definitive diagnosis of HNPCC can only be established by showing a germline mutation.[5] However, it is very important to remember that failure to identify a clearly pathogenic mutation in *hMSH2* or *hMLH1* in a person or family meeting a set of clinical criteria for HNPCC should not result in changes in clinical management decisions for at risk family members.[20] Some mutations in known genes may escape detection by the testing methods used, or an alteration in a different gene may later be found in the kindred.

ACKNOWLEDGEMENTS

This work was supported by grants from FIS 94/37, SAF95/0285, SAF96/187, SAF98-42, SAF00/81, Marató de TV3 1994 (to MAP and GC), and Fundació Catalana de Gastroenterologia. MP is a fellow of the Spanish Ministry of Education and Science.

Authors' affiliations

M Palicio, S González, Laboratori d'Investigació Gastrointestinal, Hospital de Sant Pau, Avda Sant Antoni M Claret 167, 08025 Barcelona, Spain
J Balmaña, E Marcuello, J J López López, J Brunet, Department of Medical Oncology, Hospital de Sant Pau, Avda Sant Antoni M Claret 167, 08025 Barcelona, Spain
S González, J R Germà, G Capella, Institut Català d'Oncologia, Hospital Duran I Reynals, Av Gran Vía km 2.7, 08907 L'Hospitalet, Barcelona, Spain
I Blanco, Fundació Catalana de Gastroenterologia, Barcelona, Spain
M A Peinado, G Julià, Institut de Recerca Oncològica (Department of Cancer & Metastasis), Hospital Duran I Reynals, Av Gran Vía km 2.7, 08907 L'Hospitalet, Barcelona, Spain

Correspondence to: Dr G Capellá, Laboratori de Recerca Translacional, Institut Català d'Oncologia. Av Gran Via s/n, km 2.7, 08907 L'Hospitalet, Barcelona, Spain; gcapella@ico.scs.es

REFERENCES

1 **Midgley R**, Kerr D. Colorectal cancer. *Lancet* 1999;**353**:391-9.
2 **Lynch HT**, de la Chapelle A. Genetic susceptibility to non-polyposis colorectal cancer. *J Med Genet* 1999;**36**:801-18.
3 **Vasen H**, Mecklin J, Khan P, Lynch H. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). *Dis Colon Rectum* 1991;**34**:424-5.
4 **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**:1453-6.
5 **Salovaara R**, Loukola A, Kristo P, Kaariainen H, Ahtola H, Eskelinen M, Harkonen N, Julkunen R, Kangas E, Ojala S, Tulikoura J, Valkamo E, Jarvinen H, Mecklin JP, Aaltonen LA, de la Chapelle A. Population-based molecular detection of hereditary nonpolyposis colorectal cancer. *J Clin Oncol* 2000; **18**:2193-200.
6 **Douglas FS**, O'Dair LC, Robinson M, Evans DG, Lynch SA. The accuracy of diagnoses as reported in families with cancer: a retrospective study. *J Med Genet* 1999;**36**:309-12.
7 **Tinley S**, Lynch H. Integration of family history and medical management of patients with hereditary cancers. *Cancer* 1999;**86**:1705-12.
8 **Loukola A**, de la Chapelle A, Aaltonen LA. Strategies for screening for hereditary non-polyposis colorectal cancer. *J Med Genet* 1999;**36**:819-22.
9 **Weber TK**, Conlon W, Petrelli NJ, Rodriguez-Bigas M, Keitz B, Pazik J, Farrell C, O'Malley L, Oshalim M, Abdo M, Anderson G, Stoler D, Yandell D. Genomic DNA-based hMSH2 and hMLH1 mutation screening in 32 Eastern United States hereditary nonpolyposis colorectal cancer pedigrees. *Cancer Res* 1997;**57**:3798-803.
10 **Lamberti C**, Kruse R, Ruelfs C, Caspari R, Wang Y, Jungck M, Mathiak M, Malayeri HR, Friedl W, Sauerbruch T, Propping P. Microsatellite instability: a useful diagnostic tool to select patients at high risk for hereditary non-polyposis colorectal cancer: a study in different groups of patients with colorectal cancer. *Gut* 1999;**44**:839-43.
11 **Genuardi M**, Anti M, Capozzi E, Leonardi F, Fornasarig M, Novella E, Bellacosa A, Valenti A, Gasbarrini GB, Roncucci L, Benatti P, Percesepe A, Ponz de Leon M, Coco C, de Paoli A, Valentini M, Boiocchi M, Neri G, Viel A. *MLH1* and *MSH2* constitutional mutations in colorectal cancer families not meeting the standard criteria for hereditary nonpolyposis colorectal cancer. *Int J Cancer* 1998;**75**:835-9.
12 **Chomczynski P**, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 1987;**162**:156-9.
13 **Palicio M**, Blanco I, Tortola S, Gonzalez I, Marcuello E, Brunet J, Lluis F, Gonzalez-Aguilera JJ, Peinado MA, Capella G. Intron splice acceptor site polymorphism in the *hMSH2* gene in sporadic and familial colorectal cancer. *Br J Cancer* 2000;**82**:535-7.
14 **Leach FS**, Nicolaides NC, Papadopoulos N, Liu B, Jen J, Parsons R, Peltomaki P, Sistonen P, Aaltonen LA, Nystrom-Lahti M. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell* 1993;**75**:1215-25.
15 **Peltomaki P**, Vasen HF. Mutations predisposing to hereditary nonpolyposis colorectal cancer: database and results of a collaborative study. The International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer. *Gastroenterology* 1997;**113**:1146-58.
16 **Charbonnier F**, Martin C, Scotte M, Sibert L, Moreau V, Frebourg T. Alternative splicing of hMLH1 messenger RNA in human normal cells. *Cancer Res* 1995;**55**:1839-41.
17 **Gonzalez-García I**, Moreno V, Navarro M, Marti-Rague J, Marcuello E, Benasco C, Campos O, Capella G, Peinado MA. Standardized approach for microsatellite instability detection in colorectal carcinomas. *J Natl Cancer Inst* 2000;**92**:544-9.

18 **Hamilton SR**, Liu B, Parsons RE, Papadopoulos N, Jen J, Powell SM, Krush AJ, Berk T, Cohen Z, Tetu B, Burger P, Wood P, Taqi F, Booker S, Petersen G, Offerhaus G, Tersmette A, Giardiello F, Vogelstein B, Kinzler K. The molecular basis of Turcot's syndrome. *N Engl J Med* 1995;**332**:839-47.
19 **Miyaki M**, Konishi M, Muraoka M, Kikuchi-Yanoshita R, Tanaka K, Iwama T, Mori T, Koike M, Ushio K, Chiba M. Germ line mutations of *hMSH2* and *hMLH1* genes in Japanese families with hereditary nonpolyposis colorectal cancer (HNPCC): usefulness of DNA analysis for screening and diagnosis of HNPCC patients. *J Mol Med* 1995;**73**:515-20.
20 **Wijnen J**, Khan PM, Vasen H, van der Klift H, Mulder A, van Leeuwen-Cornelisse I, Bakker B, Losekoot M, Moller P, Fodde R. Hereditary nonpolyposis colorectal cancer families not complying with the Amsterdam criteria show extremely low frequency of mismatch-repair-gene mutations. *Am J Hum Genet* 1997;**61**:329-35.
21 **Tomoda H**, Baba H, Azakawa K. Prolonged survival in hereditary nonpolyposis colorectal cancer. *Oncol Rep* 1999;**6**:321-24.
22 **Lynch HT**, Watson P, Shaw TG, Lynch JF, Harty AE, Franklin BA, Kapler CR, Tinley, Liu B, Lerman C. Clinical impact of molecular genetic diagnosis, genetic counseling, and management of hereditary cancer. *Cancer* 1999;**86**:1637-43.
23 **Aaltonen L**, Salovaara R, Kristo P, Canzian F, Hemminki A, Peltomaki P, Chadwick RB, Kaariainen H, Eskelinen M, Jarvinen H, Mecklin JP, de la Chapelle A. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. *N Engl J Med* 1998;**338**:1481-7.
24 **Liu B**, Parsons R, Papadopoulos N, Nicolaides NC, Lynch HT, Watson P, Jass JR, Dunlop M, Wyllie A, Peltomaki P, de la Chapelle A, Hamilton SR, Vogelstein B, Kinzler KW. Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients. *Nat Med* 1996;**2**:169-74.
25 **Wijnen J**, van der Klift H, Vasen H, Khan PM, Menko F, Tops C, Meijers Heijboer H, Lindhout D, Moller P, Fodde R. MSH2 genomic deletions are a frequent cause of HNPCC *Nat Genet* 1998;**20**:326-8.
26 **Liu T**, Wahlberg S, Burek E, Lindblom P, Rubio C, Lindblom A. Microsatellite instability as a predictor of a mutation in a DNA mismatch repair gene in familial colorectal cancer. *Genes Chrom Cancer* 2000;**27**:17-25.
27 **Wu Y**, Berends MJ, Post JG, Mensink RG, Verlind E, Van Der Sluis T, Kempinga C, Sijmons RH, Van Der Zee AG, Hollema H, Kleibeuker JH, Buys CH, Hofstra RM. Germline mutations of EXO1 gene in patients with hereditary nonpolyposis colorectal cancer (HNPCC) and atypical HNPCC forms. *Gastroenterology* 2001;**120**:1580-7.
28 **Tannergard P**, Lipford JR, Kolodner R, Frodin JE, Nordenskjold M, Lindblom A. Mutation screening in the hMLH1 gene in Swedish hereditary nonpolyposis colon cancer families. *Cancer Res* 1995;**55**:6092-6.
29 **Frogatt NJ**, Brassett C, Koch DJ, Evans DG, Hodgson SV, Ponder BA, Maher ER. Mutation screening of MSH2 and MLH1 mRNA in hereditary non-polyposis colon cancer syndrome. *J Med Genet* 1996;**33**:726-30.
30 **Hutter P**, Couturier A, Membrez V, Joris F, Sappino AP, Chappuis PO. Excess of hMLH1 germline mutations in Swiss families with hereditary non-polyposis colorectal cancer. *Int J Cancer* 1998;**78**:680-4.
31 **Wehner M**, Buschhausen L, Lamberti C, Kruse R, Caspari R, Propping P, Friedl W. Hereditary nonpolyposis colorectal cancer (HNPCC): eight novel germline mutations in hMSH2 and hMLH1 genes. *Hum Mut* 1997;**10**:241-4.
32 **Bubb VJ**, Curtis LJ, Cunningham C, Dunlop MG, Carothers AD, Morris RG, White S, Bird CC, Wyllie AH. Microsatellite instability and the role of hMSH2 in sporadic colorectal cancer. *Oncogene* 1996;**12**:2641-9.
33 **Kolodner RD**, Hall NR, Lipford J, Kane MF, Rao MR, Morrison P, Wirth L, Finan PJ, Burn J, Chapman P. Structure of the human MSH2 locus and analysis of two Muir-Torre kindreds for msh2 mutations. *Genomics* 1994;**24**:516-26.

7.5. Appendix 5. “What is the hereditary non-polyposis colorectal cancer syndrome?”

Judith Balmaña, Joan Brunet, Carlos Pericay, Gabriel Capellà, Miguel Ángel Peinado, Ignacio Blanco, Juan José López López, Eugenio Marcuello.

Rev Oncología 2000; 2:191-201.

What is the hereditary non-polyposis colorectal cancer syndrome?

Judith Balmaña, Joan Brunet[a], Carlos Pericay, Gabriel Capellà[b], Miguel Ángel Peinado[c], Ignacio Blanco[d], Juan José López López and Eugenio Marcuello

Servei d'Oncologia Mèdica. Hospital de la Santa Creu i Sant Pau. Barcelona. [a]Servei d'Oncologia Mèdica. Hospital Sant Joan de Reus. [b]Institut Català d'Oncologia i [c]Institut de Recerca Oncològica. L'Hospitalet de Llobregat. Barcelona. [d]Fundació Catalana de Gastroenterologia. Barcelona.

Hereditary non-polyposis colorectal cancer syndrome may account for up to 4% of the total colorectal cancer burden. The earliest reference to inheritance in colorectal cancer was a century ago and intensive research since then has resulted in the present knowledge of its molecular basis, clinical identification, and genetic diagnosis. The aim of this review is to highlight the historical evolution, genetic, molecular, and biological features that underlie the clinical management of both affected patients and at-risk family members.

Key words: Hereditary non-polyposis colorectal cancer. Mutator genes. Microsatellite instability. Genetic counseling. Surveillance.

Rev Oncología 2000; 2: 191-201.

Síndrome del cáncer de colon hereditario no polipósico (CCHNP)

El cáncer de colon hereditario no polipósico (CCHNP) es un síndrome que representa aproximadamente el 4% del total de cáncer colorrectal. La primera referencia al carácter hereditario del cáncer colorrectal fue a finales del siglo pasado, y el conocimiento actual de su base molecular, identificación clínica y diagnóstico genético es fruto de las investigaciones que se han llevado a cabo en este campo desde entonces. El objetivo de esta revisión es resaltar su evolución histórica, sus características genéticas, moleculares y biológicas que condicionan el manejo clínico de estos pacientes y de los familiares a riesgo.

Palabras clave: Cáncer de colon hereditario no polipósico. Genes mutadores. Inestabilidad de microsatélites. Consejo genético. Seguimiento.

INTRODUCTION

Colorectal cancer is the second most common cause of death from cancer and the fourth most prevalent neoplasm in western countries. Together with a better understanding of the molecular events associated with the development of colorectal cancer (CRC), there has been an increasing awareness of the role of environmental and clinical risk factors, including prior benign colonic neoplasia, long-standing extensive inflammatory bowel disease, and family history. Colonic and rectal neoplasms share many environmental risk factors and both of them are also found in individuals with specific genetic syndromes[1]. The well-recognized hereditary syndromes associated with an increased risk of colorectal cancer are familial adenomatous polyposis (FAP), the hamartomatous polyposis syndromes (Peutz-Jeghers syndrome and juvenile polyposis), and hereditary non-polyposis colorectal cancer syndrome (HNPCC). FAP represents approximately 1% of the CRC burden. Familial juvenile polyposis, Peutz-Jeghers syndrome, and inflammatory bowel disease together contribute another 1%, and it is estimated that HNPCC may account for 2% to 10% of the total colorectal cancer burden. The distinction between familial clustering of CRC and HNPCC is sometimes difficult to make, but it has very important implications for patient follow-up and counseling of at-risk family members. The purpose of this article is to review the history, molecular genetics and clinicopathological manifestations of the syndrome and to consider management options in terms of genetic counseling, screening, and surveillance in this condition.

DEFINITION AND HISTORY

HNPCC is an autosomal-dominant disorder with a marked increase in cancer susceptibility, especially

Correspondencia: Dra. J. Balmaña.
Servei d'Oncologia Mèdica, Hospital de la Santa Creu i Sant Pau.
Sant Antoni M.ª Claret, 167. 08025 Barcelona.
E-mail: jbalmana@hsp.santpau.es

Recibido el 7-5-2000.
Aceptado para su publicación el 7-6-2000.

cancer of the colorectum and endometrium. The syndrome is characterized by the early appearance of tumors, which are occasionally multiple and predominantly localized in the proximal colon. The disease was first described in 1913 by Dr. Aldred Warthin. In 1895 he noticed that his seamstress was very depressed and questioned her about her grief. She told him that she would die of cancer because most of her relatives had died of colon, gastric, or uterine cancer. In fact, she did die at a young age of endometrial carcinoma. Warthin studied her family, now known as "Warthin's Family G"[2] and documentation about her pedigree chart showed the genealogy and pathology for many relatives. This information was not fully appreciated until two extended families were described under the name of cancer family syndrome (CFS) in the mid 1960s[3] by Henry T. Lynch. These two CFS families were named Familiy "N" (Nebraska) and family "M" (Michigan) and were of particular interest because of the wide anatomic distribution of cancer sites, the large number of individuals with multiple primary cancers, the high incidence of endometrial carcinoma, and the multigenerational transmission of cancer with an autosomal dominant inheritance pattern. In 1971 Lynch published an updated review of cancer in "Warthin's Family G", which had its roots in Germany. The review was of particular interest because it demonstrated a preponderance of adenocarcinomas of the colon, endometrium, and stomach. Interestingly, gastric carcinoma was the predominant cancer in the early generations of the family but was later replaced by colon carcinoma. This change paralleled the decline of gastric carcinoma in the general population and implies an environmental influence in these hereditary cancers. During the 1970s and early 1980s geneticists and epidemiologists suggested that families being labeled as CFS were chance clusters of cancer or FAP with inadequate pathologic study. However, by the mid-1980s pedigrees consistent with CFS began to appear in the literature from around the world[4-6]. The name was changed to HNPCC to clarify the lack of multiple colonic polyps and to separate this syndrome from the polyposis syndromes. In 1984, Boland and Troncale[7] suggested the term Lynch Syndrome I to describe families with predominantly colon malignancies at an early age and Lynch Syndrome II to describe "cancer families" with colonic and extracolononic malignancies (especially those of the female genital tract). In 1989 the International Collaborative Group on HNPCC (ICG-HNPCC) was established. The aim of this group was to develop uniform diagnostic criteria for HNPCC, to improve patient and physician education about the disorder, to establish international collaborative studies and to promote the development of national HNPCC registries. In 1991, the ICG-HNPCC published what is known as the Amsterdam criteria for the clinical diagnosis of HNPCC[8]. These criteria are set out in table 1.

TABLE 1 **Criteria of the International Collaborative Group on HNPCC for the diagnosis of HNPCC (Amsterdam Criteria I)**

1. Three or more relatives with colorectal cancer, one of whom must be a first-degree relative of the other two
2. At least two affected generations
3. At least one affected individual to be aged < 50 years
4. Exclusion of familial adenomatous polyposis
5. Tumors should be verified by pathological examination

As criteria for the selection of families for research, they were originally aimed at specificity more than sensitivity. Accordingly, these criteria have proved useful in selecting families with a high likelihood of harboring a hereditary predisposition to develop cancer. Nevertheless, they have been strongly criticized because they take no account of the role of extracolonic malignancies and may be regarded as being too strict in relation to young age requirements. It is now considered that many true HNPCC families might be missed if these criteria are exclusively used in the clinical setting, leading to these families being excluded from genetic counseling, DNA testing, or surveillance. At the ninth and tenth meeting of the ICG-HNPCC, new selection requirements for collaborative studies were proposed that included extracolonic cancers associated with HNPCC. Evaluation of current literature on the tumor spectrum of HNPCC indicated that cancer of the endometrium, stomach, ovaries, small bowel, ureter, renal pelvis, brain, and hepatobiliary tract are all associated with HNPCC[9-12]. Among these tumors, cancer of the endometrium, ureter, renal pelvis, and small bowel have the highest relative risk, and are therefore the most specific for HNPCC. A set of new clinical criteria was proposed and accepted by the ICG-HNPCC in 1998 (table 2). These are known as the Amsterdam Criteria II. Moreover, several studies worldwide have shown mutations in MMR genes in families not fulfilling the Amsterdam criteria. This allowed the ICG-HNPCC to devise criteria I and II for suspected HNPCC kindreds[13]. They have the advantage that they can be applied to nuclear families and include extracolonic tumors (table 3).

HOW COMMON IS HNPCC?

A number of attempts have been made to estimate the relative burden of HNPCC cases in the overall incidence of CRC[5-6,14-19]. Estimates among all colorectal cancer patients range between 1% and 10%, most appraisals amounting to 2% to 5%. Many studies are retrospective and were performed before the Amsterdam criteria were established; some of them, although population based, may have overestimated the

TABLE 2. **Revised criteria of the International Collaborative Group on HNPCC (Amsterdam Criteria II)**

1. There should be at least three relatives with an HNPCC-associated cancer (CRC, cancer of the endometrium, small bowel, ureter, or renal pelvis)
2. One should be a first-degree relative of the other two
3. At least two successive generations should be affected
4. At least one should be diagnosed before the age of 50 years
5. Familial adenomatous polyposis should be excluded in patients with CRC, if any
6. Tumors should be verified by pathological examination

TABLE 3. **Criteria I and II for suspected HNPCC**

Criteria I. At least one item from both categories 1 and 2 should be fulfilled
- Category 1
  1. Vertical transmission of CRC
  2. At least two siblings affected with CRC in a family
- Category 2
  1. Multiple colorectal tumors (also adenomatous polyps)
  2. At least one CRC diagnosed before the age of 50 years
  3. Development of extracolonic cancer (endometrium, urinary tract, small intestine, stomach, hepatobiliary system, or ovary) in family members

Criteria II. One CRC patient with at least one of the following:
1. Early age of onset (< 40 years)
2. Endometrial, urinary tract, or small intestine cancer in the index patient or a sibling (one aged < 50 years)
3. Two siblings with other integral HNPCC extracolonic cancers (one aged < 50 years)

incidence due to geographical clustering of the syndrome in relatively small sample sizes. In 1987, Mecklin investigated the familial occurrence of cancer in a study population consisting of all CRC cases (n = 486) diagnosed under the age of 70 during a 10-year period in a single Finnish county (population: 250,000)[5]. The criteria applied for HNPCC diagnosis were three or more cases of CRC among first-degree relatives. Mecklin reported a frequency of cancer family syndrome of 3.8% of all new CRC diagnoses, a proportion which rose to 5.5% when patients with two or more first-degree relatives harboring other extracolonic malignancy were considered. In 1989, Ponz de Leon investigated the familial occurrence of cancer in 380 patients with colorectal tumors from Northern Italy[6]. In approximately 4% of all cases the familial aggregation of two or more cases of CRC among relatives, as well as the frequent localization of these tumors in the proximal colon, strongly suggested HNPCC as a possible diagnosis. A subsequent report by the same authors[16] that examined the prevalence of HNPCC using the Amsterdam criteria estimated it to represent 3.4% of all CRC cases. In Northern Ireland, Kee and Collins retrospectively investigated all patients with non-polyposis colorectal cancer before the age of 55 (n = 205) and estimated that the prevalence of HNPCC in their study might be between 1% and 2.6%[14]. The main drawback of this investigation is that the analysis was limited to the proband age under 55 years. Since early age of onset is only one of the features of HNPCC, it is possible that cancer aggregates could also be detected in many families in which the proband was older than 55 years. Furthermore, it is also possible that the incidence of familial aggregation in elderly patients could be lower. Using the Amsterdam criteria, Evans et al. performed another population-based study in the United Kingdom[17] which revealed a very low frequency (0.3%) of HNPCC. When less strict criteria were allowed, the frequency rose to 1.6%. In 1998 Aaltonen et al[18] designed a prospective study in Finland to determine the incidence of HNPCC among 509 consecutive patients with colorectal carcinoma. They screened tumor specimens for DNA replication errors, which are characteristic of hereditary colorectal cancers, and also screened for germ-line mutations of the mismatch-repair genes MLH1 and MSH2 in patients with replications errors. Their results showed that al least 2% of patients had HNPCC detected by molecular screening. Similarly, in a Spanish hospital-based study of more than 400 CRC patients, HNPCC accounted for approximately 3% of all cases (Molina M, Blanco I, Capella G: personal communication). With all these studies it seems likely that HNPCC accounts for between 0.5% and 5% of all CRC, depending on the criteria used and population studied.

GENETIC AND MOLECULAR BASIS OF HNPCC

The pattern of inheritance of HNPCC is compatible with an autosomal dominant condition with a high degree of penetrance, but the nature of the molecular genetic abnormality has only recently been clarified. When a new strand of DNA is synthesized during the process of replication, errors that are not immediately corrected by the 3' to 5' exonuclease activity of DNA polymerase are corrected by a DNA mismatch repair (MMR) system. This repair process requires the normal function of at least seven protein subunits encoded by MMR genes that are each inherited in an autosomal-dominant fashion. To date, six of these genes and proteins have been well characterized, namely hMSH2[20-22], hMLH1[23,24], hMSH6 (GTBP)[25], hPMS1 and hPMS2[26] and hMSH3[27]. Tumors lacking this repair facility have been termed RER+ (*RE*plication *E*rror prone) or MSI(+) (MicroSatellite Instability) and express a specific mutator phenotype where multiple genetic lesions accumulate, mainly in mononucleotide or dinucleotide repeats (CA)n[28,29]. Using an unbiased PCR-based DNA fingerprint technique, Ionov et al demonstrated that instability was found in

repeated Alu sequences with deletions in their poly (A) tails throughout the whole genome of colorectal carcinomas. These alterations were also present in adenomas, which supported the concept that these deletion mutations were somatic events that occurred before or early after neoplastic transformation and were therefore not the result of genetic instability of cancer cells during tumor progression. The possibility that HNPCC is caused by defects in mismatch repair genes was first suggested by Aaltonen et al[30]. Since they knew that predisposition to CRC was linked to markers on chromosome 2p16 in some families, and that other genes responsible for cancer predisposition undergo allelic loss in tumors, they searched for loss of heterozygosity at that locus in HNPCC tumors. Their results were different from those expected, since this locus was not deleted in any of the HNPCC tumors and they found shifts in the electrophoretic mobility of (CA)n dinucleotide repeat fragments of these markers, suggesting that RER had occurred in these sequences during tumorigenesis. At the same time, studies by Thibodeau et al[31] showed that there was RER+ at sites on chromosomes 5q, 15q, 17p and 18q in 28% of all the CRCs examined. This instability significantly correlated with location of the tumor in the proximal colon and, inversely, with loss of heterozygosity for chromosomes 5q, 17p, and 18q, suggesting that some CRCs may arise through a mechanism that somehow does not involve loss of heterozygosity.

Microsatellites are short repetitive sequences of DNA distributed throughout the human genome, usually dinucleotide or trinucleotide repeats, the most common one being (CA)n. They usually present non-pathological length polymorphisms and are very useful in linkage analysis. Mutation of mismatch repair-genes results in microsatellite instability, which leads to an accumulation of mutations and repeats located in coding sequences in cancer-associated genes such as TGF-β-RII[32,33], IGF-RII and BAX. These genes have been found to be inactivated in the presence of MSI through frameshift mutations at mononucleotide repeat stretches inside the coding sequence. Transforming growth factor beta (TGF-β) inhibits growth and induces apoptotic cell death in colon epithelial cells by binding to a cell surface receptor (TGF-β-RI and RII). The subtype RII gene has been found to be mutated in 90% of MSI colon cancers at the time of transition from late adenomas to overt carcinomas. BAX is a pro-apoptotic gene whose inactivation might result in the loss of the growth control exerted by the p53-programmed cell death pathway. These types of mutations are subsequently selected for because they provide selective advantage to cancer cells. However, whether mutations in p53 and APC genes are more prevalent in sporadic than in HNPCC tumors remains to be established[34-36]. Importantly, microsatellite instability is already present in the earliest lesions (adenomas) in HNPCC patients, thus indirectly supporting the notion that microsatellite instability is more likely to be a cause than a consequence of instability in cancer cells[37]. In HNPCC there is a germline mutation in one allele of any of the known MMR genes, which increases the likelihood of acquiring a second mutation. Tumorigenesis will develop when, in a given cell, a somatic mutation occurs in the remaining allele, leading to a significant deficiency in the MMR system in target tissue such as colon, endometrium, and others[38].

TABLE 4. **Bethesda Guidelines for testing of colorectal cancer for microsatellite instability**

1. Individuals with cancer in families that meet the Amsterdam Criteria
2. Individuals with two HNPCC-related cancers, including synchronous and metachronous colorectal cancers or associated extracolonic cancers (endometrial, ovarian, gastric, hepatobiliary, or small-bowel cancer or transitional cell carcinoma of the renal pelvis or ureter)
3. Individuals with CRC and a first-degree relative with CRC and/or HNPCC-related extracolonic cancer and/or a colorectal adenoma; one of the cancers diagnosed at age < 45 years, and the adenoma diagnosed at age < 40 years
4. Individuals with CRC or endometrial cancer diagnosed at age < 45 years
5. Individuals with right-sided CRC with an undifferentiated pattern (solid/cribiform) on histopathology diagnosed at age < 45 years (solid/cribiform defined as poorly differentiated or undifferentiated carcinoma composed of irregular, solid sheets of large eosinophilic cells and containing small gland-like spaces)
6. Individuals with signet-ring-cell-type CRC diagnosed at age < 45 years (composed of >Q 50% signet ring cells)
7. Individuals with adenomas diagnosed at age < 40 years

Although the MMR phenotype is found in approximately 90% of HNPCC tumors, it has also been detected in about 15% of sporadic CRC[39,40]. The Bethesda Guidelines (table 4)[41], which are based on clinical criteria, were developed to identify tumors that should be tested for RER or microsatellite instability, which could help in identifying patients who are candidates for mismatch repair gene mutation analyses, independent of the Amsterdam criteria. Initially, one of the main limitations in the use of RER was the lack of consensus about its definition. Differences in the type and number of markers analyzed, as well as in the cutoff criteria to define the microsatellite mutator phenotype have produced conflicting results[42,43]. Recently, a panel of experts[42] recommended that alterations in two or more markers from a reference panel of five loci (two mononucleotide repeats and three dinucleotide repeats) be used to identify tumors with high MSI. Alternatively, if more markers are analyzed, MSI-positive tumors should display mutations in 30% to 40% of the loci. Apart from this consensus, the BAT26 microsatellite has been proposed

as the best marker for rapid determination of the RER phenotype, because of its sensitivity, specificity, and the added advantage of not requiring normal control DNA from the same individual[44]. More recently, Gonzalez-Garcia et al[45] have provided an algorithm for the efficient characterization of MSI that could be defined by instability in two or more of four to six markers analyzed. A stepwise strategy, consisting first of a bulk screening of two loci and then a second screening of two to four additional markers, provided excellent sensitivity (≥ 97%) and specificity (100%) for MSI diagnosis.

Curiosly, MSI in sporadic colon and endometrial cancers seems to be related to mutations in MMR genes in only 10% of cases. Studies have shown that these MSI + cancers often show cytosine hypermethylation in the hMLH1 promoter region; thus epigenetic silencing of this gene may be the cause of MSI in these tumors[46,47]. Therefore, the notion that in many cases the MSI phenotype in sporadic tumors is caused by factors other than defects in any of the known MMR genes cannot be excluded, and some studies support the belief that the DNA methylator and MMR phenotype are not mutually exclusive in CRC cell lines[48].

Aaltonen et al[18] performed a study in more than 500 patients with CRC to identify the incidence of HNPCC and the feasibility of molecular screening for the disease. They identified MSI in 12% of the tumors and germ-line mutations in 2% of the patients. In addition, they used a polymerase-chain reaction (PCR)-based method to search for a common founder mutation in exon 16 of MLH1, identified in over 40 unrelated families in Finland and Sweden, which accounted for half of the germ-line mutations found in the study. A founder mutation is one arising in a previous generation that has been inherited unchanged by a proportion of all the offspring. As a conclusion these authors recommended testing for RER in patients with CRC and a family history of colorectal and endometrial tumors, and an age of less than 50 years.

Wijnen et al[49] designed a logistical model based on clinical findings to estimate the probability of detecting a germline mutation. They assessed the prevalence of MSH2 and MLH1 mutations in families suspected of having HNPCC and evaluated whether clinical findings could predict the outcome of genetic testing. They developed a strategy of molecular analysis in families suspected of having HNPCC. This strategy has been considered very effective in clinical practice since it distinguishes between kindreds with high a probability of having a mutation detected, and who are therefore candidates for direct hMHLH1 and hMSH2 mutation analysis, and kindreds with a low probability of harboring one, who are candidates, as a first step, for microsatellite analysis. We have modified and simplified this algorithm by considering only patients fulfilling the Amsterdam criteria as candidates for direct mutation analysis. In the remaining cases of high-risk CRC, MSI instability can be used as a first screening step (fig. 1). Nevertheless, our preliminary analysis of this algorithm suggests that its usefulness may be limited by the fact that MSI analysis of paraffin-embedded tissues does not always yield a conclusive result.

Moreover, the usefulness of MSI detection as an indicator of needing more extensive genetic analyses has been questioned by Furlan et al[50]. After studying 100 patients with CRC they concluded that RER screening was not a useful in identifying HNPCC since they identified only one family with germ-line mutation in the RER positive group. However, in their study the RER phenotype correlated with different clinicopathological characteristics, such as the mucinous type, intense lymphoid peritumoral infiltrates, a relatively low mitotic index, and a low incidence of lymph node metastasis, indicating that RER-positive tumors are less aggressive. These findings are in agreement with recent data reported by other authors who also correlate the RER with early-onset CRC and better survival[45, 51-53]. The study by Gryfe et al[53] confirms that sporadic CRC in young patients with high-frequency MSI is related to a significant survival advantage, independently of pathological stage and tumor grade. Moreover, the RER phenotype decreases the likelihood of lymph node and distant metastasis. These observations further support the routine use of MSI analyses since it apparently provides a novel prognostic factor in CRC.

Up to now, more than 227 different germline mutations have been described in HNPCC families[54]; these mutations are listed in the HNPCC mutation database (www.nfdht.nl). The majority of them (up to 90%) are detected in hMLH1 and hMSH2 genes, with occasional mutations in the PMS1, PMS2 and hMSH6 genes. Quite interestingly, the rate of mutations differs, depending upon the ethnic origin of the families[55-60]. The rate of mutation detection in hMSH2 and hMLH1 in HNPCC families who fulfill the Amsterdam criteria is usually around 40%-60%[56,58,61-63], although it may oscillate from 22%[64] to 86%[65]. Finally, although germline mutations in MMR genes are much less frequent in kindreds not fulfilling the Amsterdam criteria, up to 8% to 20% of families or patients featuring early-age onset of cancer, or strong clustering of CRC and other extracolonic tumors associated with Lynch syndrome (especially endometrial cancer) (table 3) harbor MMR mutations[13] and, in any case should be excluded from genetic counseling[49,62,66-68]. In our experience with the Spanish population, mutations in MMR genes have been found in 3 out of 11 families with Amsterdam criteria (27%: two in hMLH1 and one in hMSH2) and in 4 out of 21 kindreds with suspected HNPCC (19%: four in hMLH1)[69]. Our results suggest that Amsterdam criteria may be too stringent

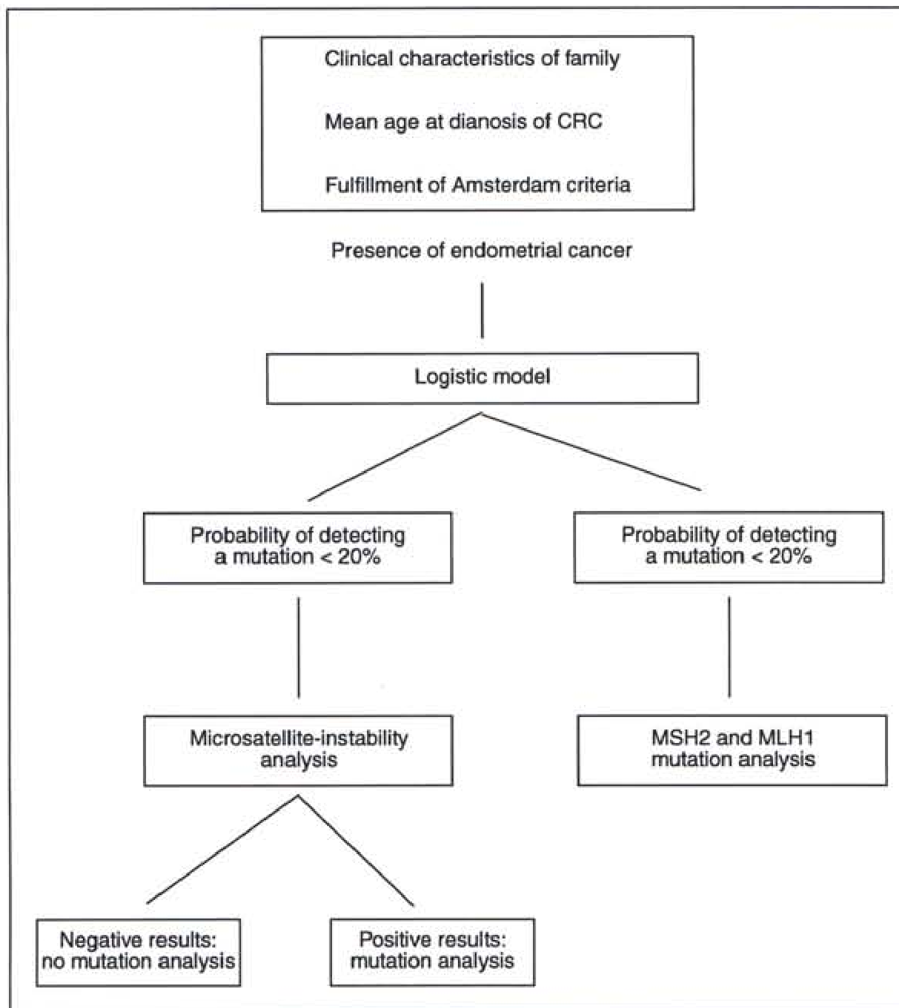


Fig. 1. Strategy of molecular analysis in families suspected of having HNPCC (adapted from Wijnen et al).

to identify all families with MMR mutations and that other genes, not yet identified, could be responsible for the increased predisposition to CRC observed in those Amsterdam (+) families not harboring MMR mutations.

CLINICAL MANIFESTATIONS

HNPCC is characterized by early-onset (usually before the age of 45-50 years) CRC with a predominance in the proximal colon[70]. Tumors tend to be poorly differentiated (solid or cribiform pattern) or with signet ring cells, and are usually of the mucinous type[71,72]. Some other typical pathological features are the Crohn-like lymphoid reaction and, apparently, an increased risk of synchronous villous adenomas[73].

The risk of a synchronous CRC in a patient with HNPCC at the time of diagnosis is approximately three times the risk in the general population (7.4% for MLH1, 6.7% for MSH2, and 2.4% for the general population)[73-75]. The rate of metachronous CRC is also higher than in the general population, with an accumulated risk of 40% during the first 10 years after surgery[76]. Such data have very important implications for index patient management at the time of cancer diagnosis and surveillance.

A study showing an overall 10-year survival advantage (69% versus 48%)[74] when comparing patients with MLH1/MSH2 mutations versus those with sporadic ones postulated that HNPCC family members with CRC have a better prognosis than patients with sporadic CRC[77-79]. Nevertheless, this apparently better survival has not been confirmed in two other studies[78,80] when adjustment for other clinical and pathological variables was performed. Alternatively, this putative better survival could be attributed to the increased representation of MSI(+) tumors in HNPCC family members.

Endometrial neoplasm is the second most common tumor[9,81] in the spectrum associated with HNPCC families. Interestingly, the cumulative risk for this tumor is higher in hMSH2 mutation carriers compared with that of hMLH1 carriers[82]. The other extracolonic types of tumor associated with HNPCC syndro-

me are cancers of the ovary, stomach, biliary tract, uro-epithelium (transitional-cell carcinoma), and kidneys (renal-cell carcinoma)[83], as well as tumors of the central nervous system in the known Turcot syndrome[84]. There is no difference in the overall risk of CRC between the MLH1 and MSH2 genotypes (84% versus 71%)[85], but rectal cancer seems to be more frequent in MSH2 kindreds[74]. Besides, risk for endometrial cancer seems to be higher than that for CRC in female patients[83,85]. These genotypic-phenotypic differences have important implications for the management of these patients, taking into account the importance of extracolonic, especially female genital-tract, screening and rectal screening in some individuals.

GENETIC COUNSELING

The compilation of a thorough family history remains the first step in the accurate diagnosis and genetic counseling in any situation in which a hereditary disorder is suspected[86,87]. Cancer risk assessment and discussion of screening and prevention options are basic aspects of preventive oncology. Nowadays, counseling should be begun prior to individual DNA testing. In this context, pretest counseling focuses on the benefits, limitations and possible adverse effects of testing, as well as on providing information about the advantages and drawbacks of screening methods. Post-test counseling sessions explore the test result, and family members' reaction to it[88].

The American Society of Clinical Oncology[89] has provided guidelines for DNA-based genetic counseling by physicians outside the research setting. Genetic testing raises medical, social, psychological, and ethical issues that must be addressed by all health care professionals, who should have adequate education in molecular genetics, pedigree construction, and the ethical and legal complexities of genetic testing. Additionally, providers of genetic risk assessment and testing services for hereditary colorectal cancer must understand the possible types of gene tests, their limitations, and their interpretation[90]. Genetic testing for HNPCC also requires provision of informed consent by the tested individual[91].

Genetic counseling in cancer should have the following aims: *a)* to educate the family about the clinical and management aspects of the disorder, its hereditary nature, the risks of cancer, the consequences of receiving gene-positive or negative results, and the recommended screening guidelines for each possible test outcome; *b)* to gain an understanding of the patient's perspective of family history and cancer experiences and the acceptability and convenience of screening regimens; and *c)* to explore the perception of risk and the anticipated meaning of any test results. Bearing these aims in mind, it seems that in our setting genetic counseling in HNPCC should be provided by a multidisciplinary team that includes oncologists, as the professionals with the greatest knowledge of the natural history of colon cancers, geneticists, surgeons, gastroenterologists, and clinical psychologists.

Because of the implications of genetic testing, results should be disclosed in person to the individual tested. For persons who test gene positive, a follow-up session is recommended in which the physician determines whether referral for psychological support is indicated. Psychological distress may also be observed among people who are told that they are not carriers of a deleterious gene because they feel survivor guilt. Finally, it is important to know the patient's culture, beliefs and traditions at the time of genetic counseling, since these factors influence decisions about lifetime surveillance for colorectal cancer or prophylactic colectomy in asymptomatic mutation positive individuals[92-94].

SCREENING AND FOLLOW-UP RECOMMENDATIONS

The value of screening at-risk first-degree relatives for CRC was first suggested by Love and Morrisey in 1984[95]. Shortly after, Mecklin et al[96] provided data on the prevalence of adenomas and carcinomas in asymptomatic at-risk subjects from 22 HNPCC families from Finland. Colonoscopy or flexible sigmoidoscopy/air contrast barium enema was performed in 137 asymptomatic subjects. Early-stage neoplasia was detected in 9% of the individuals and metachronous lesions were found in 35%. On this basis, other studies attempted to provide controlled data to examine the role of screening high-risk first-degree relatives of CRC patients associated with HNPCC. Jarvinen et al[97] examined outcomes in 133 asymptomatic first-degree relatives undergoing 3-yearly screening colonoscopy or flexible sigmoidoscopy plus barium enema and compared these outcomes with an age- and sex-matched control group who had refused screening. There was a reduction of 62% in the incidence of CRC among the study subjects which was presumably due to polypectomies. Mortality from CRC was zero in the screened group and was 36% in controls. The incidence of non-screened, extracolonic tumors was similar between the two groups.

Winawer et al[98] demonstrated that colonoscopic polypectomy reduced the incidence of CRC in a cohort of 1,418 patients. Later, they showed that colonoscopy performed 3 years after removal of adenomatous polyps detects clinically relevant lesions as effectively as follow-up examination after both 1 and 3 years[99]. If we take into account that the adenoma-carcinoma sequence is more rapid in HNPCC, a 3-year interval between colonoscopies is probably too long for at-risk

TABLE 5. **Follow-up recommendations in at-risk patients of HNPCC families**

1. Full colonoscopy to the caecum every 2 years beginning at age 20-25 years until 40 years, and then annually
2. Annual screening for endometrial cancer (endometrial aspirate or transvaginal ultrasound) beginning at age 25-35 years

individuals; therefore we think a 1-2 year interval may be recommended[100,101]. Given the known early onset of CRC in HNPCC, consensus now exists that screening of at-risk individuals (i.e. putative gene carriers on the basis of clinical history or confirmed carriers by gene studies) should begin at 20-25 years of age and should consist of complete colonoscopy every 1-2 years until the age of 40; thereafter, screening should be performed annually[102,103]. Moreover, a cost-effectiveness analysis of colorectal screening of HNPCC gene carriers[104] showed that annual surveillance of gene carriers led to an increase in life expectancy of 8 years, with a decrease in the risk of developing CRC to 20%, thus making the costs of CRC surveillance lower than those of non-surveillance.

The role of screening for extracolonic malignancy in at-risk members of HNPCC families remains unclear. It has been suggested that the high risk of endometrial cancer argues in favor of beginning screening from the age 35 years[102] by gynecological examination (endovaginal US) and endometrial aspiration, if needed. Neoplasms of the stomach and urinary tract, if present in the family, could also be screened by means of upper endoscopy and urine cytology, although no strategy can currently be firmly recommended due to the lack of good data regarding efficacy (table 5).

The role of prophylactic colectomy for gene carriers is also controversial. Some authors argue in favor of prophylactic surgery among HNPCC germline mutation carriers since a lifetime colon cancer risk of 80%-85% is fairly high[105]. On the other hand, Syngal et al106 designed a mathematical model of prevention strategy in male gene carriers and concluded that surveillance with colonoscopy is better in quality of life adjusted to life-expectancy than prophylactic surgery. Clearly, patients require genetic counseling regarding this issue.

Finally, as far as the surgical management of CRC in gene carriers is concerned, the significant incidence of synchronous and metachronous tumors in these patients has led to the recommendation that subtotal or total colectomy and ileo-sigmoid or ileorectal anastomosis be performed as primary surgical management[102,107]. However, the risk of rectal cancer after total colectomy and ileorectal anastomosis is about 12% at 12 years, which suggests that aggressive endoscopic surveillance of the rectum should be performed after abdominal colectomy[108].

In conclusion, more HNPCC research is needed in the field of available surveillance and prophylactic options among these high-risk individuals, taking into account the expected functional outcome and morbidities associated with each strategy.

References

1. Scott R, Müller H. Familial and genetic aspects of colorectal carcinogenesis. Eur J Cancer 1993; 29A: 2163-2167.
2. Warthin A. Heredity with reference to carcinoma. Arch Intern Med 1913; 12: 546-555.
3. Lynch H, Shaw M, Magnuson CW, Larsen AL, Krush AJ. Hereditary factors in cancer. Study of two large midwestern kindreds. Arch Intern Med 1966; 117: 206-212.
4. Mecklin J, Jarvinen H, Peltokallio P. Cancer family syndrome. Genetic analysis of 22 Finnish kindreds. Gastroenterology 1986; 90: 328-333.
5. Mecklin J-P. Frequency of hereditary colorectal carcinoma. Gastroenterology 1987; 93: 1021-1025.
6. Ponz de Leon M, Sassatelli R, Sacchetti C, Zanghieri G, Scalmat A, Roncucci L. Familial aggregation of tumors in the three-year experience of a population-based colorectal cancer registry. Cancer Res 1989; 49: 4344-4348.
7. Boland C, Troncale F Familial colonic cancer without antecedent polyposis. Ann Intern Med 1984; 100: 700-701.
8. Vasen H, Mecklin J, Khan P, Lynch H. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Dis Colon Rectum 1991; 34: 424-425.
9. Watson P, Lynch H Extracolonic Cancer in Hereditary Nonpolyposis Colorectal Cancer. Cancer 1993; 71: 677-685.
10. Vasen H, Wijnen J, Menko F et al. Cancer risk in families with hereditary nonpolyposis colorectal cancer (HNPCC). Int J Cancer 1996; 65: 422-425.
11. Vasen HF, Sanders EA, Taal BG et al. The risk of brain tumours in hereditary non-polyposis colorectal cancer (HNPCC). Int J Cancer 1996; 65: 422-425.
12. Aarnio M, Mecklin JP, Aaltonen LA, Nystrom-Lahti M, Jarvinen HJ. Life-time risk of different cancers in hereditary non-polyposis colorectal cancer (HNPCC) syndrome. Int J Cancer 1995; 64: 430-433.
13. Park J-G, Vasen HF, Park KJ et al. Suspected hereditary nonpolyposis colorectal cancer. International collaborative group on hereditary nonpolyposis colorectal cancer (ICG-HNPCC) criteria and results of genetic diagnosis. Dis Colon Rectum 1999; 42: 710-716.
14. Kee F, Collins BJ. How prevalent is cancer family syndrome? Gut 1991; 32: 509-512.
15. Kee F, Collins BJ. Families at risk of colorectal cancer who are they? Gut 1992; 33: 787-790.
16. Ponz de Leon M, Sassatelli R, Benatti P, Roncucci L. Identification of hereditary nonpolyposis colorectal cancer in the general population. Cancer 1993; 71: 3493-3501.
17. Evans DG, Walsh S, Jeacock J et al. Incidence of hereditary non-polyposis colorectal cancer in a population-based study of 1137 consecutive cases of colorectal cancer. Br J Surg 1997; 84: 1281-1285.

18. Aaltonen L, Salovaara R, Kristo P et al. Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. N Engl J Med 1998; 338: 1481-1487.
19. Ponz de Leon M, Pedroni M, Benatti P et al. Hereditary colorectal cancer in the general populationfrom cancer registration to molecular diagnosis. Gut 1999; 45: 32-38.
20. Peltomaki P, Aaltonen LA, Sistonen P et al. Genetic mapping of a locus predisposing to human colorectal cancer. Science 1993; 260: 810-812.
21. Fishel R, Lescoe M, Rao M et al. The human mutator gene homologue MSH2 and its association with hereditary nonpolyposis colon cancer. Cell 1993; 75: 1027-1038.
22. Leach F, Nicolaides N, Papadopoulos N et al. Mutations of a mutS homologue in hereditary nonpolyposis colorectal cancer. Cell 1993; 75: 1215-1225.
23. Bronner C, Baker S, Morrison P et al. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. Nature 1994; 36: 258-261.
24. Papadopoulos N, Nicolaides N, Wei Y-F et al. Mutation of a mutL homologue in hereditary colon cancer. Science 1994; 263: 1625-1629.
25. Palombo F, Gallinari P, Iaccarino I. GTBP, a 160-kilodalton protein essential for mismatch binding activity in human cells. Science 1995; 268: 1912-1914.
26. Nicolaides N, Papadopoulos N, Liu B et al. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. Nature 1994; 371: 75-80.
27. New L, Liu K, Crouse GF. The yeast gene MSH3 defines a new class of eukaryotic MutS homologues. Mol Gen Genet 1993; 239: 97-108.
28. Ionov Y, Peinado M, Malkbosyan S, Shibata D, Perucho M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature 1993; 363: 558-561.
29. Perucho M, Pérez-Piteira J, Soto JL, Schwartz S, Gascon A, Yamamoto H. The discovery of the microsatellite mutator phenotype and the distinction of the mutator and suppressor pathways for colorectal cancer. Rev Oncologia 1999; 1: 6-12.
30. Aaltonen L, Peltomaki P, Leach F et al. Clues to the pathogenesis of familial colorectal cancer. Science 1993; 260: 812-815.
31. Thibodeau S, Bren G, Schaid D. Microsatellite instability in cancer of the proximal colon. Science 1993; 260: 816-819.
32. Markowitz S, Wang J, Myeroff L et al. Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. Science 1995; 268: 1336-1338.
33. Parsons R, Myeroff L, Liu B et al. Microsatellite instability and mutations of the transforming growth factor beta type II receptor gene in colorectal cancer. Cancer Res 1995; 55: 5548-5550.
34. Lazar V, Grandjouan S, Bognel C et al. Accumulation of multiple mutations in tumour suppressor genes during colorectal tumorigenesis in HNPCC patients. Hum Mol Genet 1994; 3: 2257-2260.
35. Losi L, Ponz de Leon M, Jiricny J et al. K-ras and p53 mutations in hereditary non-polyposis colorectal cancers. Int J Cancer 1997; 74: 94-96.
36. Huang J, Papadopoulos N, McKinley AJ. APC mutations in colorectal tumors with mismatch repair deficiency. Proc Natl Acad Sci USA 1996; 93: 9049-9054.
37. Shibata D, Peinado MA, Ionov Y, Malkoshyan S, Perucho M. Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. Nat Genet 1994; 6: 273-281.
38. Peltomaki P, Lothe R, Aaltonen R. Microsatellite instability is associated with tumors that characterize the hereditary non-polyposis colorectal carcinoma syndrome. Cancer Res 1993; 53: 5853-5855.
39. Lothe R, Peltomaki P, Meling G et al. Genomic instability in colorectal cancer relationship to clinicopathological variables and family history. Cancer Res 1993; 53: 5849-5852.
40. Pedroni M, Tamassia MG, Percesepe A et al. Microsatellite instability in multiple colorectal tumors. Int J Cancer 1999; 81: 1-5.
41. Rodriguez-Bigas M, Boland C, Hamilton S et al. A National Cancer Institute workshop on HNPCC syndrome meeting highlights and Bethesda guidelines. J Natl Cancer Inst 1997; 89: 1762-1785.
42. Boland CR, Thibodeau SN, Hamilton SR D, JR, et al. 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.
43. Perucho M. 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 (letter to the editor) Cancer Res 1998, 58: 5248-5257.
44. Zhou XP, Hoang JM, Li YJ et al. Determination of the replication error phenotype in human tumors without the requirement for matching normal DNA by analysis of mononucleotide repeat microsatellites. Genes Chromosomes Cancer 1998; 21: 101-107.
45. González-García I, Moreno V, Navarro M et al. Standardized Approach for microsatellite Instability detection in colorectal carcinomas. J Natl Cancer Inst 2000; 92: 544-549.
46. Kane MF, Loda M, Gaida GM et al. Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cells. Cancer Res 1997; 57: 808-811.
47. Esteller M, Levine R, Baylin SB, Ellenson LH, Herman JG. MLH1 promoter hypermethylation is associated with the microsatellite instability phenotype in sporadic endometrial carcinomas. Oncogene 1998; 16: 2413-2417.
48. Pao MM, Liang G, Tsai YC, Xiong Z, Laird PW, Jones PA. DNA methylator and mismatch repair phenotypes are not mutually exclusive in colorectal cancer cell lines. Oncogene 2000; 19: 943-952.
49. Wijnen J, Vasen H, Meera Khan P et al. Clinical findings with implications for genetic testing in families with clustering of colorectal cancer. N Engl J Med 1998; 339: 511-518.
50. Furlan D, Tibiletti M, Taborelli M, Albarello L, Cornaggia M, Capella C. The value of microsatellite instability in the detection of HNPCC families and of sporadic colorectal cancers with special biological features An investigation on a series of 100 consecutive cases. Ann Oncol 1998; 9: 901-906.
51. Capozzi E, Della Puppa L, Fornasarig M, Pedroni M, Boiocchi M, Viel A. Evaluation of the replication error phenotype in relation to molecular and clinicopathological features in hereditary and early onset colorectal cancer. Eur J Cancer 1999; 35: 289-295.
52. Johannsdottir JT, Bergthorsson JT, Gretarsdottir S et al. Replication error in colorectal carcinoma association with loss of heterozygosity at mismatch repair loci and clinicopathological variables. Cancer Res 1999; 19: 1821-1826.

53. Gryfe R, Kim H, Hsieh ETK et al. Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer. N Engl J Med 2000; 342: 69-77.
54. Peltomaki P, Vasen HF. ICG-HNPCC Mutations predisposing to hereditary nonpolyposis colorectal cancer database and results of a collaborative study. The International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer. Gastroenterology 1997; 113: 1146-1158.
55. Miyaki M, Konishi M, Muraoka M et al. Germ line mutations of hMSH2 and hMLH1 genes in Japanese families with hereditary nonpolyposis colorectal cancer (HNPCC) usefulness of DNA analysis for screening and diagnosis of HNPCC patients. J Mol Med 1995; 73: 515-520.
56. Viel A, Genuardi M, Capozzi E et al. Characterization of MSH2 and MLH1 mutations in Italian families with hereditary nonpolyposis colorectal cancer. Genes Chromosomes Cancer 1997; 18: 8-18.
57. Kolodner RD, Tytell JD, Schmeits JL et al. Germ-line MSH6 mutations in colorectal cancer families. Cancer Res 1999; 59: 5068-5074.
58. Wang Q, Lasset C, Desseigne F et al. Prevalence of germline mutations of hMLH1, hMSH2, hPMS1, hPMS2, and hMSH6 in 75 French kindreds with nonpolyposis colorectal cancer. Hum Genet 1999; 105: 79-85.
59. Bapat BV, Madlensky L, Temple LKF et al. Family history characteristics, tumor microsatellite instability and germline MSH2 and MLH1 mutations in hereditary colorectal cancer. Hum Genet 1999; 104: 167-176.
60. Planck M, Koul A, Fernebro E et al. hMLH1, hMSH2 and hMSH6 mutations in hereditary non-polyposis colorectal cancer families from Southern Sweden. Int J Cancer 1999; 83: 197-202.
61. Liu B, Parsons R, Papadopoulos N et al. Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients [see comments]. Nat Med 1996; 2: 169-174.
62. Wijnen J, Khan PM, Vasen H et al. Hereditary nonpolyposis colorectal cancer families not complying with the Amsterdam criteria show extremely low frequency of mismatch-repair-gene mutations. Am J Hum Genet 1997; 61: 329-335.
63. Bai Y-Q, Akiyama Y, Nagasaki H et al. Predominant germ-line mutation of the hMSH2 gene in Japanese hereditary non-polyposis colorectal cancer kindreds. Int J Cancer 1999; 82: 512-515.
64. Wijnen J, Vasen H, Khan PM et al. Seven new mutations in hMSH2, an HNPCC gene identified by denaturing gradient-gel electrophoresis. Am J Hum Genet 1995; 56: 1060-1066.
65. Nystrom-Lahti M, Wu Y, Moisio AL et al. DNA mismatch repair gene mutations in 55 kindreds with verified or putative hereditary non-polyposis colorectal cancer. Hum Mol Genet 1996; 5: 763-769.
66. Genuardi M, Anti M, Capozzi E et al. MLH1 and MSH2 constitutional mutations in colorectal cancer families not meeting the standard criteria for hereditary nonpolyposis colorectal cancer. Int J Cancer 1998; 75: 835-839.
67. Park J-G, Park YJ, Wijnen JT, Vasen HFA. Gene-environment interaction in hereditary nonpolyposis colorectal cancer with implications for diagnosis and genetic testing. Int J Cancer 1999; 82: 516-519.
68. Heinimann K, Scott RJ, Buerstedde J-M et al. Influence of selection criteria on mutation detection in patients with hereditary nonpolyposis colorectal cancer. Cancer 1999; 85: 2512-2518.
69. Balmaña J, Brunet J, Palicio M et al. hMSH2 and hMLH1 gene mutations in Spanish families with colorectal cancer (CRC) clustering. High prevalence of mutations in families not meeting Amsterdam criteria (AC). Proc Am Soc Clin Oncol 2000; 19 (609): 2399.
70. Fitzgibbons R, Lynch H, Stanislav G et al. Recognition and treatment of patients with hereditary colon cancer (Lynch syndromes I and II). Ann Surg 1987; 206: 289-295.
71. Vasen H, Den Hartog Jager F, Menko F, Nagengast F. Screening for hereditary non-polyposis colorectal cancer a study of 22 kindreds in The Netherlands. Am J Med 1989; 86: 278-281.
72. Mecklin J, Sipponen P, Jarvinen H Histopathology of colorectal carcinomas and adenomas in cancer family syndrome. Dis Colon Rectum 1986; 29: 849-853.
73. Jass J, Smyrk T, Stewart S, Lane M, Lanspa S, Lynch H. Pathology of hereditary non-polyposis colorectal cancer. Anticancer Res 1994; 14: 1631-1634.
74. Lin KM, Shashidharan M, Ternent CA et al. Colorectal and extracolonic cancer variations in MLH1/MSH2 hereditary nonpolyposis colorectal cancer kindreds and the general population. Dis Colon Rectum 1998; 41: 428-433.
75. Myrhoj T, Bisgaard ML, Bernstein I, Svendsen LB, Sondergaard JO, Bulow S. Hereditary non-polyposis colorectal cancer clinical features and survival. Results from the Danish HNPCC register. Scand J Gastroenterol 1997; 32: 572-576.
76. Lynch H, Watson P, Lanspa S. Natural History of colorectal cancer in hereditary non-polyposis colorectal cancer (Lynch syndromes I and II). Dis Colon Rectum 1988; 31: 439-444.
77. Sankila R, Aaltonen LA, Jarvinen HJ, Mecklin JP. Better survival rates in patients with MLH1-associated hereditary colorectal cancer [see comments]. Gastroenterology 1996; 110: 682-687.
78. Percesepe A, Benatti P, Roncucci L et al. Survival analysis in families affected by hereditary non-polyposis colorectal cancer. Int J Cancer 1997; 71: 373-376.
79. Tomoda H, Baba H, Azakawa K. Prolonged survival in hereditary nonpolyposis colorectal cancer. Oncol Rep 1999; 6: 321-324.
80. Bertario L, Russo A, Sala P et al. Survival of patients with hereditary colorectal cancer comparison of HNPCC and colorectal cancer in FAP patients with sporadic colorectal cancer. Int J Cancer 1999; 80: 183-187.
81. Vasen H, Watson P, Mecklin J et al. The Epidemiology of Endometrial Cancer in Hereditary Nonpolyposis Colorectal Cancer. Cancer Res 1994; 14: 1675-1678.
82. Jager AC, Bisgaard ML, Myrhoj T, Bernstein I, Rehfeld JF, Nielsen FC. Reduced frequency of extracolonic cancers in hereditary nonpolyposis colorectal cancer families with monoallelic hMLH1 expression. Am J Hum Genet 1997; 61: 129-138.
83. Aarnio M, Sankila R, Pukkala E et al. Cancer risk in mutation carriers of DNA-mismatch-repair genes. Int J Cancer 1999; 81: 214-218.
84. Hamilton SR, Liu B, Parsons RE et al. The molecular basis of Turcot's syndrome. N Engl J Med 1995; 332: 839-847.
85. Lin KM, Shashidharan M, Thorson AG et al. 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.
86. Tinley S, Lynch H. Integration of family history and medical management of patients with hereditary cancers. Cancer 1999; 86: 1705-1712.
87. Balmaña J, Brunet J, Capellà G et al. Identificación de dos familias con cáncer de colon hereditario no polipósico (CCHNP) y criterios de Amsterdam. Relevancia del árbol genealógico y las medidas de seguimiento. Med Clin (Barc) 2000; 114: 56-59.

88. Menko FH, Wijnen JT, Khan PM, Vasen HF, Oosterwijk MH. Genetic counseling in hereditary nonpolyposis colorectal cancer. Oncology 1996; 10: 71-76.
89. Statement of the American Society of Clinical Oncology Genetic testing for cancer susceptibility. J Clin Oncol 1996; 14: 1730-1736.
90. Giardiello F, Brensinger J, Petersen G et al. The use and interpretation of commercial APC gene testing for familial adenomatous polyposis. N Engl J Med 1997; 336: 823-827.
91. Geller G, Botkin J, Green M et al. Genetic testing for susceptibility to adult-onset cancer the process and content of informed consent. JAMA 1997; 277: 1467-1474.
92. Lynch HT, Drouhard T, Vasen HF et al. Genetic counseling in a Navajo hereditary nonpolyposis colorectal cancer kindred. Cancer 1996; 77: 30-35.
93. Lynch HT, Lemon S, Smyrk T et al. Genetic counseling in hereditary nonpolyposis colorectal cancer an extended family with MSH2 mutation. Am J Gastroenterol 1996; 91: 2489-2493.
94. Lynch HT, Watson P, Shaw TG et al. Clinical impact of molecular genetic diagnosis, genetic counseling, and management of hereditary cancer. Cancer 1999; 86: 1637-1643.
95. Love R, Morrisey J. Colonoscopy in asymptomatic individuals with a family history of colorectal cancer. Arch Int Med 1984; 144: 2209-2212.
96. Mecklin JP, Jarvinen HJ, Aukee S, Elomaa I, Karjalainen K. Screening for colorectal carcinoma in cancer family syndrome kindreds. Scan J Gastroenterol 1987; 22: 449-453.
97. Jarvinen HJ, Mecklin JP, Sistonen P. Screening reduces colorectal cancer rate in families with hereditary nonpolyposis colorectal cancer [see comments]. Gastroenterology 1995; 108: 1405-1411.
98. Winawer SJ, Zauber AG, Nah Ho M et al. Prevention of colorectal cancer by colonoscopic polypectomy. N Engl J Med 1993; 329: 1977-1981.
99. Winawer S, Zauber A, O'Brien M et al. Randomized comparison of surveillance intervals after colonoscopic removal of newly diagnosed adenomatous polyps. N Engl J Med 1993; 328: 901-906.
100. Vasen H, Mecklin J, Watson P et al. Surveillance in hereditary nonpolyposis colorectal cancer an international cooperative study of 165 families. Dis Colon Rectum 1993; 36: 1-4.
101. Vasen HF, Nagengast FM, Khan PM. Interval cancers in hereditary non-polyposis colorectal cancer (Lynch syndrome) [letter]. Lancet 1995; 345: 1183-1184.
102. Burke W, Petersen G, Lynch P et al. Recommendations for follow-up care of individuals with an inherited predisposition to cancer. JAMA 1997; 277: 915-919.
103. Winawer S, Fletcher R, Miller L et al. Colorectal cancer screening clinical guidelines and rationale. Gastroenterology 1997; 112: 594-642.
104. 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; 82: 1632-1637.
105. Church JM. Prophylactic colectomy in patients with hereditary nonpolyposis colorectal cancer. Ann Med 1996; 28: 479-482.
106. Syngal S, Weeks JC, Schrag D, Garber JE, Kuntz KM. Benefits of colonoscopic surveillance and prophylactic colectomy in patients with hereditary nonpolyposis colorectal cancer mutations. Ann Int Med 1998; 129: 787-796.
107. Mecklin J, Jarvinen H. Treatment and Follow-Up Strategies in Hereditary Nonpolyposis Colorectal Carcinoma. Dis Colon Rectum 1993; 36: 927-929.
108. Rodriguez-Bigas MA, Vasen HF, Pekka-Mecklin J et al. Rectal cancer risk in hereditary nonpolyposis colorectal cancer after abdominal colectomy. International Collaborative Group on HNPCC. Ann Surg 1997; 225: 202-207.

8 References

1: Warthin, A (1913) "Heredity with reference to carcinoma" Arch Intern Med 12: 546-555.

2: Lynch, H.T., Shaw, M.W., Magnuson, C.W., Larsen, A.L., Krush, A.J. (1966) "Hereditary factors in cancer. Study of two large midwestern kindreds." Arch Intern Med 117 (2): 206-212.

3: Mecklin, J.P, Jarvinen, H.J., Peltokallio,P. (1986) "Cancer family syndrome. Genetic analysis of 22 Finnish kindreds" Gastroenterology 90 (2): 328-233.

4: Mecklin, J.P (1987) "Frequency of hereditary colorectal carcinoma" Gastroenterology 93 (5): 1021-1025.

5: Ponz de Leon, M, Sassatelli, R, Sacchetti, C., Zanghieri, G., Scalmati, A., Roncucci, L. (1989) "Familial aggregation of tumors in the three-year experience of a population-based colorectal cancer registry" Cancer Res 49 (15): 4344-4348.

6: Boland, C.R. and Troncale, F.J. (1984) "Familial colonic cancer without antecedent polyposis" Ann Intern Med 100 (5): 700-701.

7: Vasen, H.F, Mecklin, J.P., Khan, P.M., Lynch, H.T. (1991) "The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC)" Dis Colon Rectum 34 (5): 424-425.

8: Watson, P and Lynch, H.T. (1993) "Extracolonic cancer in hereditary nonpolyposis colorectal cancer" Cancer 71 (3): 677-85.

9: Vasen, H.F, Sanders, E.A., Sanders, E.A., Taal, B.G., Nagengast, F.M., Griffioen, G., Menko, F.H., Kleibeuker, J.H., Houwing-Duistermaat, J.J., Meera Khan, P. (1996) "The risk of brain tumours in hereditary non-polyposis colorectal cancer (HNPCC)" Int J Cancer 65 (4): 422-425.

10: Aarnio, M, Mecklin, J.P, Aaltonen, L.A, Nystrom-Lahti, M, Jarvinen, H.J. (1995) "Life-time risk of different cancers in hereditary non-polyposis colorectal cancer (HNPCC) syndrome" Int J Cancer 64 (6): 430-433.

11: Boland, C.R, Thibodeau, S.N, Hamilton, S.R, Sidransky, D, Eshleman, J.R., Burt, R.W., Meltzer, S.J., Rodriguez-Bigas, M. ., Fodde, R., Ranzani, G.N., Srivastava, S. (1998) "A National Cancer Institute Workshop on Micro Satellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer" Cancer Res 58 (22): 5248-5257.

12: Ionov, Y., Peinado, M.A, Malkhosyan, S., Shibata, D., Perucho, M. (1993) "Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis" Nature 363 (6429): 558-561.

13: Umar, A., Boland, C.R., Terdiman, J.P., Syngal, S., de la Chapelle, A., Ruschoff, J., Fishel, R., Lindor, N.M., Burgart, L.J., Hamelin, R., Hamilton, S.R., Hiatt, R.A., Jass, J., Lindblom, A., Lynch, H.T., Peltomaki, P., Ramsey, S.D., Rodriguez-Bigas, M.A., Vasen, H.F., Hawk, E.T., Barrett, J.C., Freedman, A.N., Srivastava, S. (2004) “Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability” J Natl Cancer Inst 96 (4): 261-268.

14: Lynch, H T. and A. de la Chapelle (2003) Hereditary colorectal cancer” N Engl J Med 348 (10): 919-32.

15: Nystrom-Lahti, M., Wu, Y., Moisio, A.L., Hofstra, R.M., Osinga, J., Mecklin, J.P., Jarvinen, H.J., Leisti, J., Buys, C.H., de la Chapelle, A., Peltomaki, P. (1996) “DNA mismatch repair gene mutations in 55 kindreds with verified or putative hereditary non-polyposis colorectal cancer” Hum Mol Genet 5 (6): 763-769.

16: Evans, D.G., Walsh, S., Jeacock, J., Robinson, C., Hadfield, L., Davies, D.R., Kingston, R. (1997) “Incidence of hereditary non-polyposis colorectal cancer in a population-based study of 1137 consecutive cases of colorectal cancer” Br J Surg 84 (9): 1281-5.

17: Aaltonen, L.A., Salovaara, R., Kristo, P., Canzian, F., Hemminki, A., Peltomaki, P., Chadwick, R.B., Kaariainen, H., Eskelinen, M., Jarvinen, H., Mecklin, J.P., de la Chapelle, A. (1998) “Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease” N Engl J Med 338 (21): 1481-1487.

18: Planck, M., Koul, A., Fernebro, E., Borg, A., Kristoffersson, U., Olsson, H., Wenngren, E., Mangell, P., Nilbert, M. (1999) “hMLH1, hMSH2 and hMSH6 mutations in hereditary non-polyposis colorectal cancer families from southern Sweden” Int J Cancer 83 (2): 197-202.

19: Cunningham, J.M., Kim, C.Y., Christensen, E.R., Tester, D.J., Parc, Y., Burgart, L. J., Halling, K.C., McDonnell, S.K., Schaid, D.J., Walsh, C., Kubly, V., Nelson, H., Michels, V.V., Thibodeau, S.N. (2001) “The frequency of hereditary defective mismatch repair in a prospective series of unselected colorectal carcinomas” Am J Hum Genet 69 (4): 780-790.

20: Caldes, T., Godino, J., de la Hoya, M., Garcia Carbonero, I., Perez Segura, P., Eng, C., Benito, M., Diaz-Rubio, E. (2002) “Prevalence of germline mutations of MLH1 and MSH2 in hereditary nonpolyposis colorectal cancer families from Spain” Int J Cancer 98 (5): 774-779.

21: Gille, J.J., Hogervorst, F.B., Pals, G., Wijnen, J.T., Van Schooten, R.J., Dommering, C. J., Meijer, G. A., Craanen, M.E., Nederlof, P.M., de Jong, D., McElgunn, C.J., Schouten, J.P., Menko, F.H. (2002) “Genomic deletions of MSH2 and MLH1 in colorectal cancer families detected by a novel mutation detection approach” Br J Cancer 87 (8): 892-897.

22: Katballe, N., Christensen, M., Wikman, F. P., Orntoft, T. F., Laurberg, S., (2002) “Frequency of hereditary non-polyposis colorectal cancer in Danish colorectal cancer patients” Gut 50 (1): 43-51.

23: Cederquist, K., Emanuelsson, M., Goransson, I., Holinski-Feder, E., Muller-Koch, Y., Golovleva, I., Gronberg, H. (2004) “Mutation analysis of the MLH1, MSH2 and MSH6 genes in patients with double primary cancers of the colorectum and the endometrium: a population-based study in northern Sweden” Int J Cancer 109 (3): 370-376.

24: de la Chapelle, A. (2005) “The incidence of Lynch syndrome” Fam Cancer 4 (3): 233-237.

25: Hampel, H., Frankel, W. L., Martin, E., Arnold, M., Khanduja, K., Kuebler, P., Nakagawa, H., Sotamaa, K., Prior, T. W., Westman, J., Panescu, J., Fix, D., Lockman, J., Comeras, I., de la Chapelle, A. (2005) “Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer)” N Engl J Med 352 (18): 1851-60.

26: Ollikainen, M., Abdel-Rahman, W. M., Moisio, A. L., Lindroos, A., Kariola, R., Jarvela, I., Poyhonen, M., Butzow, R., Peltomaki, P. (2005) “Molecular analysis of familial endometrial carcinoma: a manifestation of hereditary nonpolyposis colorectal cancer or a separate syndrome?” J Clin Oncol 23 (21): 4609-4616.

27: Pinol, V., Andreu, M., Castells, A., Paya, A., Bessa, X., Rodrigo, J. (2004) “Frequency of hereditary non-polyposis colorectal cancer and other colorectal cancer familial forms in Spain: a multicentre, prospective, nationwide study” Eur J Gastroenterol Hepatol 16 (1): 39-45.

28: Pinol, V., Castells, A., Andreu, M., Castellvi-Bel, S., Alenda, C., Llor, X., Xicola, R. M., Rodriguez-Moranta, F., Paya, A., Jover, R., Bessa, X. (2005)“Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer” JAMA 293 (16): 1986-1994.

29: Fishel, R., Lescoe, M.K., Rao, M.R., Copeland, N.G., Jenkins, N.A., Garber, J., Kane, M., Kolodner, R. (1993) “The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer” Cell 75 (5): 1027-1038.

30: Bronner, C.E., Baker, S. M., Morrison, P.T., Warren, G., Smith, L. G., Lescoe, M. K., Kane, M., Earabino, C., Lipford, J., Lindblom, A. et al. (1994) “Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer” Nature 368 (6468): 258-261.

31: Papadopoulos, N., Nicolaides, N. C., Wei, Y. F., Ruben, S. M., Carter, K. C., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D., Fraser, C. M., Adams, M. D. et al., (1994) “Mutation of a mutL homolog in hereditary colon cancer” Science 263 (5153): 1625-9.

32: Palombo, F., Gallinari, P., Iaccarino, I., Lettieri, T., Hughes, M., D'Arrigo, A., Truong, O., Hsuan, J. J., Jiricny, J. (1995) “GTBP, a 160-kilodalton protein essential for mismatch-binding activity in human cells” Science 268(5219): 1912-1914.

33: Nicolaides, N. C., Papadopoulos, N., Liu, B., Wei, Y. F., Carter, K. C., Ruben, S. M., Rosen, C. A., Haseltine, W. A., Fleischmann, R. D., Fraser, C. M., et al., (1994) “Mutations of two PMS homologues in hereditary nonpolyposis colon cancer” Nature 371 (6492): 75-80.

34: Strand, M., Earley, M. C., Crouse, G. F., Petes, T. D. (1995) “Mutations in the MSH3 gene preferentially lead to deletions within tracts of simple repetitive DNA in Saccharomyces cerevisiae” Proc Natl Acad Sci U S A 92 (22): 10418-21.

35: Aaltonen, L.A., P.Peltomaki, et al. (1993) “Clues to the pathogenesis of familial colorectal cancer” Science 260 (5109): 812-6.

36: Peltomaki, P., Lothe, R. A., Aaltonen, L. A., Pylkkanen, L., Nystrom-Lahti, M., Seruca, R., David, L., Holm, R., Ryberg, D., Haugen, A. et al, (1993) “Microsatellite instability is associated with tumors that characterize the hereditary non-polyposis colorectal carcinoma síndrome” Cancer Res 53 (24): 5853-5.

37: Shibata, D., Peinado, M. A., Ionov, Y., Malkhosyan, S., Perucho, M. (1994) “Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation” Nat Genet 6(3): 273-81.

38: Reitmair, A. H., Cai, J. C., Bjerknes, M., Redston, M., Cheng, H., Pind, M. T., Hay, K., Mitri, A., Bapat, B. V., Mak, T. W., Gallinger, S. (1996) “MSH2 deficiency contributes to accelerated APC-mediated intestinal tumorigenesis” Cancer Res 56 (13): 2922-6.

39: Lothe, R.A., Peltomaki, P., Meling, G.I., Aaltonen, L.A., Nystrom-Lahti, M., Pylkkanen, L., Heimdal, K., Andersen, T.I., Moller, P., Rognum, T.O. et al, (1993) “Genomic instability in colorectal cancer: relationship to clinicopathological variables and family history” Cancer Res 53 (24): 5849-52.

40: Pedroni, M., Tamassia, M.G., Percesepe, A., Roncucci, L., Benatti, P., Lanza, G., Jr., Gafa, R., Di Gregorio, C., Fante, R., Losi, L., Gallinari, L., Scorcioni, F., Vaccina, F., Rossi, G., Cesinaro, A. M., Ponz de Leon, M. (1999) “Microsatellite instability in multiple colorectal tumors” Int J Cancer 81 (1): 1-5.

41: Salovaara, R., Loukola, A., Kristo, P., Kaariainen, H., Ahtola, H., Eskelinen, M., Harkonen, N., Julkunen, R., Kangas, E., Ojala, S., Tulikoura, J., Valkamo, E., Jarvinen, H., Mecklin, J. P., Aaltonen, L. A., de la Chapelle, A. (2000) “Population-based molecular detection of hereditary nonpolyposis colorectal cancer” J Clin Oncol 18 (11): 2193-2200.

42: Kane, M. F., Loda, M., Gaida, G.M., Lipman, J., Mishra, R., Goldman, H., Jessup, J.M., Kolodner, R. (1997) “Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines” Cancer Res 53 (24): 5849-5852.

43: McGivern, A., Wynter, C.V., Whitehall, V.L., Kambara, T., Spring, K.J., Walsh, M. D., Barker, M.A., Arnold, S., Simms, L.A., Leggett, B.A., Young, J., Jass, J.R. (2004) "Promoter hypermethylation frequency and BRAF mutations distinguish hereditary non-polyposis colon cancer from sporadic MSI-H colon cancer" Fam Cancer 3 (2): 101-107.

44: Domingo, E., Niessen, R. C., Oliveira, C., Alhopuro, P., Moutinho, C., Espin, E., Armengol, M., Sijmons, R. H., Kleibeuker, J. H., Seruca, R., Aaltonen, L. A., Imai, K., Yamamoto, H., Schwartz, S., Jr., Hofstra, R. M. (2005) "BRAF-V600E is not involved in the colorectal tumorigenesis of HNPCC in patients with functional MLH1 and MSH2 genes" Oncogene 24 (24): 3995-3998.

45: Bessa, X., Balleste, B., Andreu, M., Castells, A., Bellosillo, B., Balaguer, F. Castellvi - Bel, S., Paya, A., Jover, R., Alenda, C., Tito, L., Martinez-Villacampa, M., Vilella, A., Xicola, R. M., Pons, E., Llor, X. (2008) "A prospective, multicenter, population-based study of BRAF mutational analysis for Lynch syndrome screening" Clin Gastroenterol Hepatol 6 (2): 206-214.

46: Thibodeau, SN, French, AJ, Roche, PC, Cunningham, JM, Tester, DJ, Lindor, N M, Moslein, G, Baker, SM, Liskay, RM, Burgart, L J, Honchel, R, Halling, KC (1996) "Altered expression of hMSH2 and hMLH1 in tumors with microsatellite instability and genetic alterations in mismatch repair genes" Cancer Res 56 (21): 4836-4840.

47: Casey, G., Lindor, N.M., Papadopoulos, N., Thibodeau, S.N., Moskow, J., Steelman, S., Buzin, C.H., Sommer, S.S., Collins, C.E., Butz, M., Aronson, M. Gallinger, S., Barker, M.A., Young, J.P., Jass, J.R., Hopper, J.L., Diep, A., Bapat, B., Salem, M., Seminara, D., Haile, R. (2005) "Conversion analysis for mutation detection in MLH1 and MSH2 in patients with colorectal cancer" JAMA 293 (7): 799-809.

48: Caldes, T., Godino, J., Sanchez, A., Corbacho, C., De la Hoya, M., Lopez, Asenjo, J., Saez, C., Sanz, J., Benito, M., Ramon, Y. Cajal S., Diaz-Rubio, E. (2004) "Immunohistochemistry and microsatellite instability testing for selecting MLH1, MSH2 and MSH6 mutation carriers in hereditary non-polyposis colorectal cancer" Oncol Rep 12 (3): 621-629.

49: Fitzgibbons, R. J., Jr., Lynch, H. T., Stanislav, G. V., Watson, P. A., Lanspa, S. J., Marcus, J. N., Smyrk, T., Kriegler, M. D., Lynch, J. F. (1987) "Recognition and treatment of patients with hereditary nonpolyposis colon cancer (Lynch syndromes I and II)" Ann Surg 206 (3): 289-295.

50: Lynch, H.T., Lynch, J. (2000) "Lynch syndrome: genetics, natural history, genetic counseling, and prevention" J Clin Oncol 18 (21 Suppl): 19S-31S.

51: Mecklin, J. P., Sipponen, P., Jarvinen, H.J. (1986) „Histopathology of colorectal carcinomas and adenomas in cancer family syndrome" Dis Colon Rectum 29 (12): 849-853.

52: Jass, J. R., Smyrk, T. C., Stewart, S. M., Lane, M. R., Lanspa, S. J., Lynch, H. T. (1994) “Pathology of hereditary non-polyposis colorectal cancer” Anticancer Res 14 (4B): 1631-1634.

53: Jenkins, M. A., Hayashi, S., O'Shea, A. M., Burgart, L. J., Smyrk, T. C., Shimizu, D., Waring, P. M., Ruszkiewicz, A. R., Pollett, A. F., Redston, M., Barker, M. A., Baron, J. A., Casey, G. R., Dowty, J. G., Giles, G. G., Limburg, P., Newcomb, P., Young, J. P., Walsh, M. D., Thibodeau, S. N., Lindor, N. M., Lemarchand, L., Gallinger, S., Haile, R. W., Potter, J. D., Hopper, J. L., Jass, J. R. (2007) “Pathology features in Bethesda guidelines predict colorectal cancer microsatellite instability: a population-based study” Gastroenterology 133 (1): 48-56.

54: Lindgren, G., Liljegren, A., Jaramillo, E., Rubio, C., Lindblom, A. (2002) “Adenoma prevalence and cancer risk in familial non-polyposis colorectal cancer” Gut 50 (2): 228-234.

55: Lin, K. M., Shashidharan, M., Ternent, C.A., Thorson, A.G., Blatchford, G.J., Christensen, M.A., Lanspa, S.J., Lemon, S.J., Watson, P., Lynch, H.T. (1998) “Colorectal and extracolonic cancer variations in MLH1/MSH2 hereditary nonpolyposis colorectal cancer kindreds and the general population” Dis Colon Rectum 41(4): 428-433.

56: Lynch, H. T., Lynch, J. F. (2005): “What the physician needs to know about Lynch syndrome: an update” Oncology (Williston Park) 19(4): 455-463.

57: Giraldez, M.D., Castellvi-Bel, S., Balaguer, F., Gonzalo, V., Ocana, T., Castells, A. (2008) “Lynch syndrome in colorectal cancer patients” Expert Rev Anticancer Ther 8 (4): 573-583.

58: Aarnio, M., Sankila, R., Pukkala, E., Salovaara, R., Aaltonen, L. A., de la Chapelle, A., Peltomaki, P., Mecklin, J. P., Jarvinen, H. J. (1999) “Cancer risk in mutation carriers of DNA-mismatch-repair genes” Int J Cancer 81 (2): 214-218.

59: Lu, H.K., Broaddus, R.R. (2005) “Gynecologic Cancers in Lynch Syndrome/HNPCC” Fam Cancer 4 (3): 249-254.

60: Lu, K.H., Dinh, M., Kohlmann, W., Watson, P., Green, J., Syngal, S., Bandipalliam, P., Chen, L. M., Allen, B., Conrad, P., Terdiman, J., Sun, C., Daniels, M., Burke, T., Gershenson, D. M., Lynch, H., Lynch, P., Broaddus, R. R. (2005) “Gynecologic cancer as a "sentinel cancer" for women with hereditary nonpolyposis colorectal cancer syndrome” Obstet Gynecol 105(3): 569-574.

61: Lu, K.H., Schorge, J.O., Rodabaugh, K.J., Daniels, M.S., Sun, C.C., Soliman, P.T., White, K.G., Luthra, R., Gershenson, D.M., Broaddus, R.R. (2007) “Prospective determination of prevalence of lynch syndrome in young women with endometrial cancer” J Clin Oncol 25(33): 5158-5164.

62: Dunlop, M.G., Farrington, S.M., Carothers, A.D., Wyllie, A.H., Sharp, L., Burn, J., Liu, B., Kinzler, K.W., Vogelstein, B. (1997) “Cancer risk associated with germline DNA mismatch repair gene mutations” Hum Mol Genet 6(1): 105-110.

63: Hamilton, S.R., Liu, B., Parsons, R.E., Papadopoulos, N., Jen, J., Powell, S.M., Krush, A.J., Berk, T., Cohen, Z., Tetu, B. et al. (1195) “The molecular basis of Turcot's syndrome” N Engl J Med 332 (13): 839-847.

64: Chan, T., Yuen, S., Chung, L., Ho, J., Kwan, K., Fan, Y., Chan, A., Leung, S. (1999) “Germline hMSH2 and differential somatic mutations in patients with Turcot's syndrome. Genes Chromosomes Cancer 25:75-81,

65: De Rosa, M., Fasano, C., Panariello, L., Scarano, M. I., Belli, G., Iannelli, A., Ciciliano, F., Izzo, P. (2000) “Evidence for a recessive inheritance of Turcot's syndrome caused by compound heterozygous mutations within the PMS2 gene” Oncogene 19 (13): 1719-23.

66: Goecke, T., Schulmann, K., Engel, C., Holinski-Feder, E., Pagenstecher, C., Schackert, H. K., Kloor, M., Kunstmann, E., Vogelsang, H., Keller, G., Dietmaier, W., Mangold, E., Friedrichs, N., Propping, P., Kruger, S., Gebert, J., Schmiegel, W., Rueschoff, J., Loeffler, M., Moeslein, G. (2006) “Genotype-phenotype comparison of German MLH1 and MSH2 mutation carriers clinically affected with Lynch syndrome: a report by the German HNPCC Consortium” J Clin Oncol 24(26): 4285-92.

67: Peltomaki, P., Gao, X., Mecklin, J. P. (2001) “Genotype and phenotype in hereditary nonpolyposis colon cancer: a study of families with different vs. shared predisposing mutations” Fam Cancer 1(1): 9-15.

68: Bandipalliam, P., Garber, J., Kolodner, R. D., Syngal, S. (2004) “Clinical presentation correlates with the type of mismatch repair gene involved in hereditary nonpolyposis colon cancer” Gastroenterology 126(3):936-937.

69: Kastrinos, F., Stoffel, E. M., Balmana, J., Steyerberg, E. W., Mercado, R., Syngal, S. (2008) “Phenotype comparison of MLH1 and MSH2 mutation carriers in a cohort of 1,914 individuals undergoing clinical genetic testing in the United Status” Cancer Epidemiol Biomarkers Prev 17(8): 2044-2051.

70: Vasen, H. F., Stormorken, A., Menko, F. H., Nagengast, F. M., Kleibeuker, J. H., Griffioen, G., Taal, B. G., Moller, P., Wijnen, J. T. (2001) “MSH2 mutation carriers are at higher risk of cancer than MLH1 mutation carriers: a study of hereditary nonpolyposis colorectal cancer familias” J Clin Oncol 19(20): 4074-4080.

71: Plaschke, J., Engel, C., Kruger, S., Holinski-Feder, E., Pagenstecher, C., Mangold, E., Moeslein, G., Schulmann, K., Gebert, J., von Knebel Doeberitz, M., Ruschoff, J., Loeffler, M., Schackert, H. K. (2004) “Lower incidence of colorectal cancer and later age of disease onset in 27 families with pathogenic MSH6 germline mutations compared with families with MLH1 or MSH2 mutations: the German Hereditary Nonpolyposis Colorectal Cancer Consortium” J Clin Oncol 22(22): 4486-94.

72: Watson, P., Vasen, H. F., Mecklin, J. P., Bernstein, I., Aarnio, M., Jarvinen, H. J., Myrhoj, T., Sunde, L., Wijnen, J. T., Lynch, H. T. (2008) “The risk of extra-colonic, extra-endometrial cancer in the Lynch syndrome” Int J Cancer 123 (2): 444-449.

73: Hendriks, Y. M., Wagner, A., Morreau, H., Menko, F., Stormorken, A., Quehenberger, F., Sandkuijl, L., Moller, P., Genuardi, M., Van Houwelingen, H., Tops, C., Van Puijenbroek, M., Verkuijlen, P., Kenter, G., Van Mil, A., Meijers, Heijboer, H., Tan, G. B., Breuning, M. H., Fodde, R., Wijnen, J. T., Brocker-Vriends, A. H., Vasen, H. (2004) “Cancer risk in hereditary nonpolyposis colorectal cancer due to MSH6 mutations: impact on counseling and surveillance” Gastroenterology 127(1): 17-25.

74: Cederquist, K., Emanuelsson, M., Wiklund, F., Golovleva, I., Palmqvist, R., Gronberg, H. (2005) “Two Swedish founder MSH6 mutations, one nonsense and one missense, conferring high cumulative risk of Lynch syndrome” Clin Genet 68(6): 533-541.

75: Senter, L., Clendenning, M., Sotamaa, K., Hampel, H., Green, J., Potter, J. D., Lindblom, A., Lagerstedt, K., Thibodeau, S. N., Lindor, N. M., Young, J., Winship, I., Dowty, J.G., White, D.M., Hopper, J.L., Baglietto, L., Jenkins, M.A., de la Chapelle, A. (2008) “The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations” Gastroenterology 135(2): 419-28.

76: South, C.D., Hampel, H., Comeras, I., Westman, J.A., Frankel, W.L., de la Chapelle, A. (2008) “The frequency of Muir-Torre syndrome among Lynch syndrome families” J Natl Cancer Inst 100 (4): 277-281.

77: Jarvinen, H. J., Aarnio, M., Mustonen, H., Aktan-Collan, K., Aaltonen, L. A., Peltomaki, P., De La Chapelle, A., Mecklin, J. P. (2000) “Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer” Gastroenterology 118(5): 829-34.

78: Mecklin, J. P., Jarvinen, H. J. (2005)”Surveillance in Lynch syndrome” Fam Cancer “4(3): 267-71.

79: Stoffel, E.M., Turgeon, D.K., Stockwell, D.H., Zhao, L., Normolle, D.P., Tuck, M. K., Bresalier, R.S., Marcon, N.E., Baron, J.A., Ruffin, M.T., Brenner, D.E., Syngal, S. (2008) “Missed adenomas during colonoscopic surveillance in individuals with Lynch Syndrome (hereditary nonpolyposis colorectal cancer)” Cancer Prev Res 1(6): 470-475.

80: Stoffel, E.M., Turgeon, D.K., Stockwell, D.H., Normolle, D.P., Tuck, M.K., Marcon, N.E., Baron, J.A., Bresalier, R.S., Arber, N., Ruffin, M.T., Syngal, S., Brenner, D.E. (2008) ”Chromoendoscopy detects more adenomas than colonoscopy using intensive inspection without dye spraying” Cancer Prev Res 1(7): 507-513.

81: Boks, D E., Trujillo, A.P., Voogd, A.C., Morreau, H., Kenter, G.G., Vasen, H.F. (2002) “Survival analysis of endometrial carcinoma associated with hereditary nonpolyposis colorectal cancer” Int J Cancer 102(2): 198-200.

82: Broaddus, R.R., Lynch, H.T., Chen, L.M., Daniels, M.S., Conrad, P., Munsell, M.F., White, K.G., Luthra, R., Lu, K.H. (2006) “Pathologic features of endometrial carcinoma associated with HNPCC: a comparison with sporadic endometrial carcinoma” Cancer 106(1): 87-94.

83: Vasen, H.F., Moslein, G., Alonso, A., Bernstein, I., Bertario, L., Blanco, I., Burn, J., Capella, G., Engel, C., Frayling, I., Friedl, W. ,Hes, F.J. ,Hodgson, S., Mecklin, J.P., Moller, P., Nagengast, F., Parc, Y., Renkonen-Sinisalo, L., Sampson, J.R., Stormorken, A. ,Wijnen, J. (2007) "Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer)" J Med Genet 44(6): 353-362.

84: Schmeler, K.M., Lynch, H.T., Chen, L.M., Munsell, M.F., Soliman, P.T., Clark, M.B., Daniels, M.S., White, K.G., Boyd-Rogers, S.G., Conrad, P.G., Yang, K.Y., Rubin, M.M., Sun, C.C., Slomovitz, B.M., Gershenson, D.M., Lu, K.H. (2006) "Prophylactic surgery to reduce the risk of gynecologic cancers in the Lynch syndrome" N Engl J Med 354(3): 261-269.

85: Syngal, S., Fox, E.A., Eng, C., Kolodner, R.D., Garber, J.E. (2000) "Sensitivity and specificity of clinical criteria for hereditary non-polyposis colorectal cancer associated mutations in MSH2 and MLH1" J Med Genet 37(9): 641-645.

86: Kievit, W., de Bruin, J.H., Adang, E.M., Ligtenberg, M.J., Nagengast, F.M., van Krieken, J.H., Hoogerbrugge, N. (2004) "Current clinical selection strategies for identification of hereditary non-polyposis colorectal cancer families are inadequate: a meta-analysis" Clin Genet 65(4): 308-16.

87: Grover, S., Stoffel, E.M., Bussone, L., Tschoegl, E., Syngal, S. (2004) "Physician assessment of family cancer history and referral for genetic evaluation in colorectal cancer patients" Clin Gastroenterol Hepatol 2(9): 813-819.

88: Wijnen, J.T., Vasen, H.F., Khan, P.M., Zwinderman, A.H., van der Klift, H., Mulder, A., Tops, C., Moller, P., Fodde, R. (1998) "Clinical findings with implications for genetic testing in families with clustering of colorectal cancer" N Engl J Med 339(8): 511-518.

89: Lipton, L. R., Johnson, V., Cummings, C., Fisher, S., Risby, P., Eftekhar, Sadat, A. T., Cranston, T., Izatt, L., Sasieni, P., Hodgson, S.V., Thomas, H.J., Tomlinson, I.P. (2004) "Refining the Amsterdam Criteria and Bethesda Guidelines: testing algorithms for the prediction of mismatch repair mutation status in the familial cancer clinic" J Clin Oncol 22(24): 4934-4943.

90: Barnetson, R.A., Tenesa, A., Farrington, S.M., Nicholl, I.D., Cetnarskyj, R., Porteous, M.E., Campbell, H., Dunlop, M.G. (2006) "Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer" N Engl J Med 354(26): 2751-2763.

91: Chen, S., Wang, W., Lee, S., Nafa, K., Lee, J., Romans, K., Watson, P., Gruber, S. B., Euhus, D., Kinzler, K.W., Jass, J., Gallinger, S., Lindor, N. M., Casey, G., Ellis, N., Giardiello, F.M., Offit, K., Parmigiani, G. (2006) "Prediction of germline mutations and cancer risk in the Lynch syndrome" JAMA 296(12): 1479-1487.

92: Marroni, F., Pastrello, C., Benatti, P., Torrini, M., Barana, D., Cordisco, E. L., Viel, A., Mareni, C., Oliani, C., Genuardi, M., Bailey-Wilson, J. E., Ponz de Leon, M., Presciuttini, S. (2006) "A genetic model for determining MSH2 and MLH1 carrier

probabilities based on family history and tumor microsatellite instability" Clin Genet 69(3): 254-262.

93: Steyerberg, E.W., Balmana, J., Stockwell, D.H., Syngal, S. (2007) "Data reduction for prediction: A case study on robust coding of age and family history for the risk of having a genetic mutation" Stat Med 26(30): 5545-5556.

94: Castells, A., Andreu, M. (2007) [EPICOLON project: contribution to the knowledge of Lynch syndrome and other familial or hereditary colorectal cancer] Med Clin (Barc) 128(2): 55-60.

95: Balaguer, F., Balmana, J., Castellvi-Bel, S., Steyerberg, E.W., Andreu, M., Llor, X., Jover, R., Syngal, S., Castells, A. (2007) "Validation and Extension of the PREMM(1,2) Model in a Population-Based Cohort of Colorectal Cancer Patients" Gastroenterology 134:39-46.