

# Risk of colon cancer in hereditary non-polyposis colorectal cancer patients as predicted by fuzzy modeling: Influence of smoking

Rhonda M Brand, David D Jones, Henry T Lynch, Randall E Brand, Patrice Watson, Ramesh Ashwathnayan, Hemant K Roy

Rhonda M Brand, Randall E Brand, Hemant K Roy, Department of Internal Medicine, Evanston Northwestern Healthcare and Feinberg School of Medicine at Northwestern University, Evanston IL, United States

David D Jones, Department of Biological Systems Engineering, University of Nebraska-Lincoln, Lincoln NE, United States

Henry T Lynch, Patrice Watson, Ramesh Ashwathnayan, Hereditary Cancer Institute, Creighton University, United States  
Supported by a grant from the American College of Gastroenterology

Correspondence to: Dr. Rhonda M Brand, Division of Emergency Medicine and Department of Internal Medicine, Evanston Northwestern Healthcare and Feinberg School of Medicine at Northwestern University, SB 521 Walgreen's Building, 2650 Ridge Ave, Evanston, IL 60201, United States. rhbrand@enh.org  
Telephone: +1-847-5704074 Fax: +1-847-7335451

Received: 2006-03-02 Accepted: 2006-03-20

may enable formulation of clinical risk scores, thereby allowing individualization of CRC prevention strategies.

© 2006 The WJG Press. All rights reserved.

**Key words:** Hereditary non-polyposis colorectal cancer; Lynch syndrome; Smoking; Colorectal cancer; Fuzzy modeling; Risk assessment

Brand RM, Jones DD, Lynch HT, Brand RE, Watson P, Ashwathnayan R, Roy HK. Risk of colon cancer in hereditary non-polyposis colorectal cancer patients as predicted by fuzzy modeling: Influence of smoking. *World J Gastroenterol* 2006; 12(28): 4485-4491

<http://www.wjgnet.com/1007-9327/12/4485.asp>

## Abstract

**AIM:** To investigate whether a fuzzy logic model could predict colorectal cancer (CRC) risk engendered by smoking in hereditary non-polyposis colorectal cancer (HNPCC) patients.

**METHODS:** Three hundred and forty HNPCC mismatch repair (MMR) mutation carriers from the Creighton University Hereditary Cancer Institute Registry were selected for modeling. Age-dependent curves were generated to elucidate the joint effects between gene mutation (hMLH1 or hMSH2), gender, and smoking status on the probability of developing CRC.

**RESULTS:** Smoking significantly increased CRC risk in male hMSH2 mutation carriers ( $P < 0.05$ ). hMLH1 mutations augmented CRC risk relative to hMSH2 mutation carriers for males ( $P < 0.05$ ). Males had a significantly higher risk of CRC than females for hMLH1 non smokers ( $P < 0.05$ ), hMLH1 smokers ( $P < 0.1$ ) and hMSH2 smokers ( $P < 0.1$ ). Smoking promoted CRC in a dose-dependent manner in hMSH2 in males ( $P < 0.05$ ). Females with hMSH2 mutations and both sexes with the hMLH1 groups only demonstrated a smoking effect after an extensive smoking history ( $P < 0.05$ ).

**CONCLUSION:** CRC promotion by smoking in HNPCC patients is dependent on gene mutation, gender and age. These data demonstrate that fuzzy modeling

## INTRODUCTION

Accurate risk-stratification is essential for combating the 50000 yearly deaths from colorectal cancer (CRC) in the United States<sup>[1]</sup>. The best-established risk factor is a familial predisposition to CRC, which is implicated in one-quarter of all CRC cases<sup>[2]</sup>. While determining a family history can be readily accomplished, risk quantification which is critical for tailoring screening strategies, remains remarkably imprecise. For instance, even in documented carriers of CRC predisposing genes, clinical expression can be quite varied due to modulation by numerous confounding endogenous and exogenous variables<sup>[3,4]</sup>.

Hereditary non-polyposis colorectal cancer (HNPCC) represents a case in point. This autosomal dominant condition is the most common cancer predisposing syndrome engendering a  $> 70\%$  lifetime risk of developing CRC<sup>[5]</sup>. Furthermore, we have recently demonstrated that cigarette smoking of a male carrying hMLH1 mutations (versus hMSH2) increases the hazard of CRC by 1.4-, 1.6- and 2.0-fold respectively<sup>[6]</sup>. However, optimal management strategies (colonoscopic surveillance versus prophylactic colectomy) are unclear secondary to characteristic phenotypic heterogeneity, i.e. marked variations in age of onset of cancers<sup>[7]</sup>. Thus, HNPCC represents an excellent paradigm to study the gene-environment joint effect hypothesis.

Incorporating these important findings into clinical

practice is hindered by the inability to accurately quantitate the risk modulation engendered by the joint effects of genetics and environmental factors. Moreover, the inadequacy of conventional statistical approaches to model the complex nature of many of the CRC risk factors further limits application of these data to patient management. While landmark studies have explored the age of CRC diagnosis in HNPCC<sup>[8]</sup>, these estimates along with others in the literature have not yet factored in the genetic/environmental influences that determine the phenotypic heterogeneity. The focus of our study was on this phenotypic heterogeneity which is of major importance to clinicians who care for these challenging patients. We believe that by knowing the genetic and environmental risk factors, we can individualize more accurately the risk analysis, which, to our knowledge, has not been previously reported.

One approach from the engineering literature that has recently received attention for cancer risk assessment is fuzzy logic. This powerful modeling technique has been successfully used for pattern recognition and image processing and its unique ability to transcend the typical black or white approaches in standard modeling and to capture the “shades of gray” has great promise for clinical medicine<sup>[9]</sup>. While typical statistical approaches function well when the data are normally distributed and values are near the mean, this approach is often inadequate at the threshold. For instance, a very high prostate specific antigen (PSA) has excellent predictive ability for prostate cancer, but the optimal clinical management of a patient with a mildly elevated value is unclear. Fuzzy logic overcomes these limitations of conventional statistics by allowing partial membership function. In our PSA example, instead of categorizing values as either normal or abnormal, a fuzzy approach would allow one to place a value as one quarter in the normal group and three-quarters in the abnormal group. Thus, through the creation of fuzzy sets, elements can have degrees of membership on a continuum (e.g. a value can be “normal, slightly elevated, moderately elevated or highly elevated”).

Another unique attribute of fuzzy modeling is that, unlike traditional models, it does not require prior knowledge of the system being modeled. It is a “model-free” form in which natural rules are developed from the data rather than imposing rules on the modeling system. The result of this “model-free” system is still a conversion from inputs to outputs, similar to traditional algorithms<sup>[10]</sup>. Another strength of fuzzy modeling lies in its ability to model data points that may be outside of traditional inclusion boundaries and thus allowing accurate modeling with less data.

Past reports have demonstrated that fuzzy modeling can improve performance characteristics of tumor markers over conventional applications<sup>[11-13]</sup>. Our previous work with conventional statistical approaches (COX proportional hazard modeling) indicated that tobacco use, gender and mutated gene play an important role in phenotypic presentation of CRC progression in HNPCC patients as a group, but lack the ability to predict an individual risk of CRC (e.g. not sensitive to dose effects or interactions of factors). We, therefore, explored the

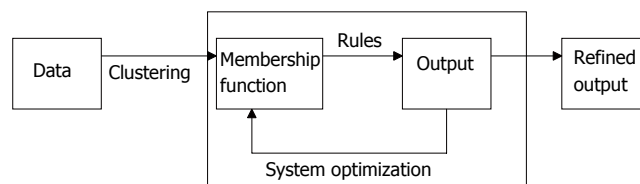


Figure 1 Overview of fuzzy methodology.

ability of fuzzy modeling to predict CRC risk in germline mutation carriers in these individual HNPCC patients by factoring the gene type, gender and tobacco use status in the present study.

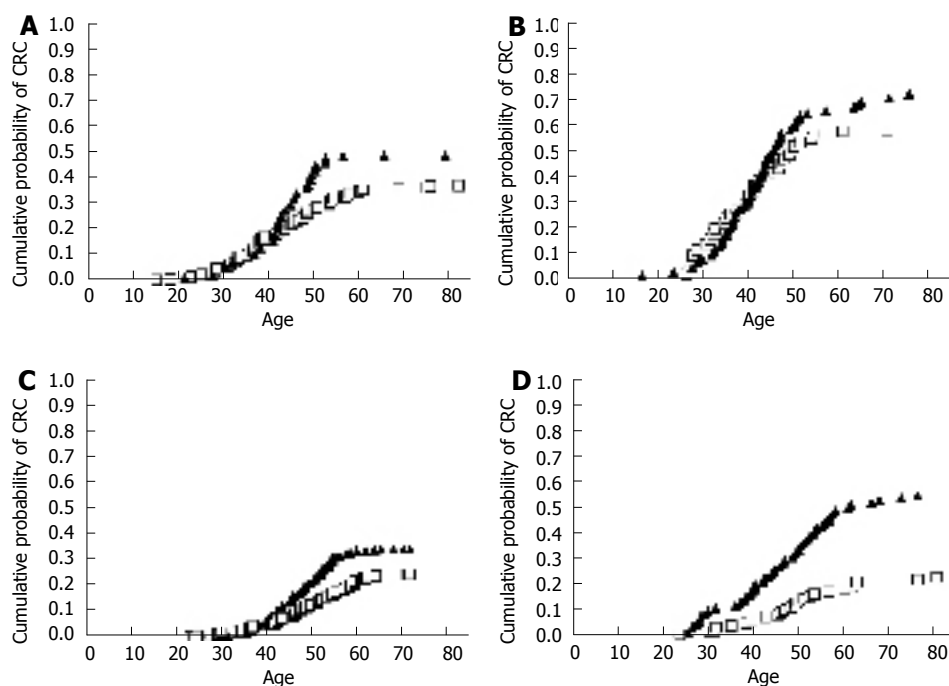
## MATERIALS AND METHODS

### Database

The Hereditary Cancer Center at Creighton University is one of the oldest and largest registries for diverse hereditary cancer syndrome, containing information on over 200 000 individuals of whom approximately 600 are verified MMR mutation carriers (Lynch Database). The database contains patient and family information, surveillance and treatment information as well as gene mutations and lifestyle data. The inclusion criteria were HNPCC as documented by either a MMR germline mutation positivity or clinical HNPCC from a patient who had a family member with a documented MMR mutation. For example, if the patient has HNPCC and his mother has a documented hMSH2 mutation, we would consider the patient to have an hMSH2 mutation. Tobacco data were obtained by self-report and family report or by abstraction from medical records. A patient was classified as a tobacco user if he/she reported ever regularly using (or was reported to have ever regularly used) cigarettes, cigars, a pipe, tobacco chew, or snuff. Five hundred and ninety-six mutation carriers were identified from 62 HNPCC families. For this analysis we only focused on cigarette smokers. Of these, 340 (60.4%) had information on tobacco use and were included in our study (158 nonsmokers and 182 smokers). In a further analysis, 271 patients (113 of 182 smokers and 158 nonsmokers) with a more detailed smoking history including calculated pack-years were selected.

### Modeling

The entire modeling procedure was performed using the MatLab Fuzzy Toolbox (Matlab, Version 6.1-Release 13, Natick, MA). Figure 1 provides an overview of the basic fuzzy modeling algorithm. The data (either categorical or continuous) were inputted into the program and a clustering algorithm led to the development of membership functions. The fuzzy clustering method used produces descriptions of each of the input vectors as belonging to one (or more) fuzzy sets with a specific membership in each of the sets. The inclusion of a continuous variable (pack-years and age) produces non-categorical (aka, fuzzy) memberships. Rules are then developed from these membership functions which successfully produce a mapping from the input space to the output space as previously described<sup>[10]</sup>. Furthermore,



**Figure 2** Scattergram demonstrating the probability of developing CRC (cumulative lifetime risk) for each subject based on smoking status,  $\blacktriangle$  = Smokers,  $\square$  = Non-Smokers. Data are divided by gene mutation and sex with (A) hMLH1: Female, (B) hMLH1: Male, (C) hMSH2: Female and (D) hMSH2: Male ( $P < 0.05$ ).

these rules represent fuzzy relationships between the variables, even if the variables themselves are categorical. The refinement of these rules is accomplished using the Adaptive Neural Fuzzy Inference System (ANFIS; 16) which acts as a feedback loop to further refine the rules until they are optimized to give the best fit to the data. Overfitting of the model to the data is not exclusively addressed. However, the clustering methods (subtractive clustering) employed in the modeling scheme tend to partition the data space in such a manner as to maximize the cluster density while simultaneously maximizing the separation of the clusters which would limit overfitting.

This modeling technique was applied to the patients selected from the Lynch database with gene mutation, sex, smoking status and age as the input and risk of developing colorectal cancer as the output. In a second study, the effect of pack years was added as an additional input. In this case, dividing a group of smokers with a given mutation, sex and age further by smoke years made the numbers in each group quite small. We therefore used the data as the training set and a theoretical set of conditions as the input to generate the model output. The models produced a cumulative risk of CRC that ranged between 0 and 1. Age and pack years were fuzzified in the program.

### Statistical analysis

Results are presented either as a scattergram of the actual model output for each patient in the database (Figures 2-4) or as an output of the model given a set of theoretical conditions (Figure 5). The statistical procedures used followed the methods described by Steel and Torrie<sup>[14]</sup>. The data for each cohort were paired and compared using a Kolmagrov-Smirnov (KS) 2-sample test. The KS test was considered to be conservative and useful when hypothesis about discrete distributions was tested. The test is motivated by the need to compare 2 independent samples and the null hypothesis is that each sample originates from

identical distributions (i.e., the data are from the same population). Critical values for the KS test are inversely proportional to the square root of the total number of observations. The nature of our data dictated that the critical values were computed using unequal populations ( $n_1 \neq n_2$ ). Furthermore, the nature of the data and results only required a comparison of this type. Other analyses, such as a test of trend or analysis of variance, were not deemed beneficial.

## RESULTS

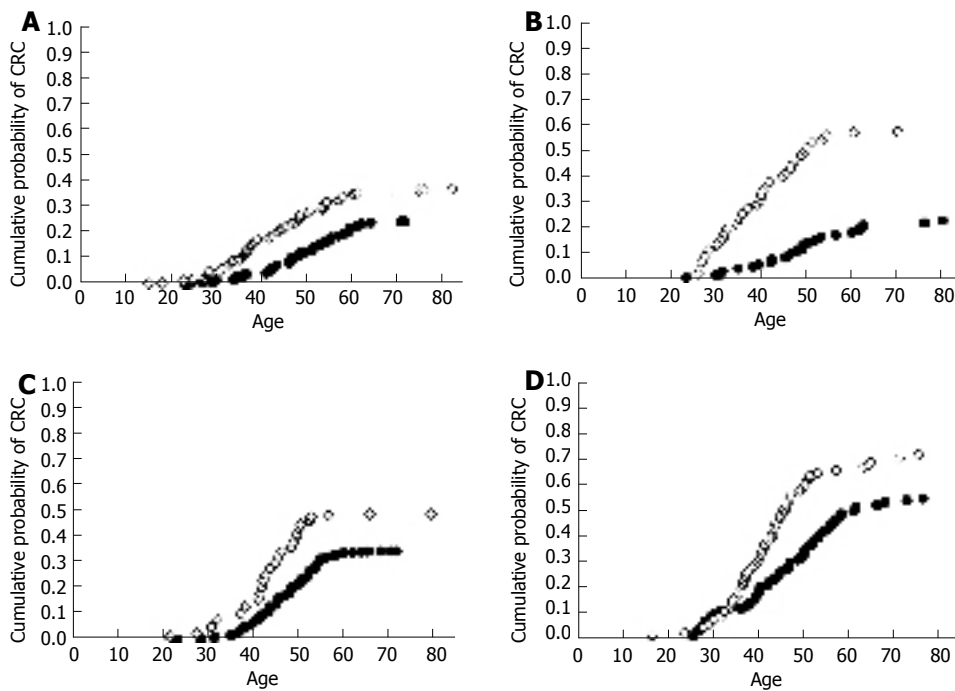
### Smoking status

The influence of genetic mutation combined with sex, smoking status and age is demonstrated in Figure 2. There was a clear impact of cigarette smoking on the age-adjusted risk of developing CRC for all conditions tested (gene mutated and gender).

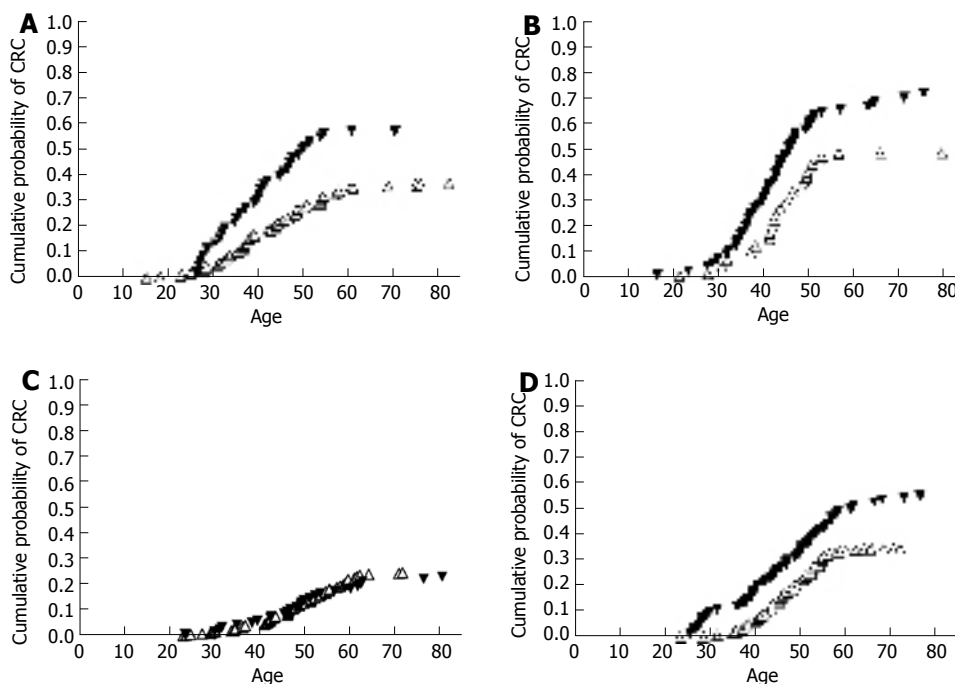
When a male patient with the hMSH2 gene mutation carrier smokes, he markedly increased his risk of developing CRC by up to 2.4-fold at the age of 78 (Figure 2D,  $P < 0.05$ ). In the case of a mutation in the hMLH1 gene, smoking increased the risk of CRC at the maximal age tested by approximately 1.3-fold for males when compared to non-smokers (Figure 2B). Females with the hMLH1 mutation showed a 1.3-fold increased risk of developing CRC and female smokers with the hMSH2 mutation had a 1.4-fold greater risk of developing CRC when compared to their non-smoking age-matched controls (Figures 2A and 2C).

### Gene

The model output was then reexamined as a function of the genetic mutation (Figure 3). For males, patients with a hMLH1 mutation had either a 2.5- or a 1.3-fold greater risk of developing CRC than those with a hMSH2 for non-smokers and smokers respectively (Figures 3B and 3D,  $P <$



**Figure 3** Scattergram demonstrating the probability of developing CRC (cumulative lifetime risk) for each subject based on genetic mutation,  $\diamond$  = hMLH1,  $\bullet$  = hMSH2. Data are divided by sex and smoking status with (A) Female: non-smokers ( $P < 0.001$ ), (B) Female: smokers ( $P < 0.001$ ), (C) Male: non-smokers ( $P < 0.001$ ) and (D) Male: smokers ( $P < 0.05$ ). The  $P$  values show the probability failing to reject the null hypothesis that the data originate from identical populations, indicating that the curves are different.



**Figure 4** Scattergram demonstrating the probability of developing CRC (cumulative lifetime risk) for each subject based on gender,  $\triangle$  = female,  $\blacktriangledown$  = male. Data are divided by gene mutation and smoking status with (A) hMLH1: non-smokers, (B) hMLH1: smokers ( $P < 0.1$ ), (C) hMSH2: non-smokers ( $P > 0.05$ ) and (D) hMSH2: smokers ( $P < 0.1$ ). The  $P$  values show the probability of failing to reject the null hypothesis that the data originate from identical populations, indicating that the curves are different.

0.05). The difference was greater for the non-smokers than the smokers because subjects with hMSH2 who smoke increased their rate of CRC greater than the non-smokers. Female non-smokers with the hMLH1 mutation showed a 1.5-fold increase in CRC risk as compared to the hMSH2 subjects, whereas it was 1.4-fold higher for smokers.

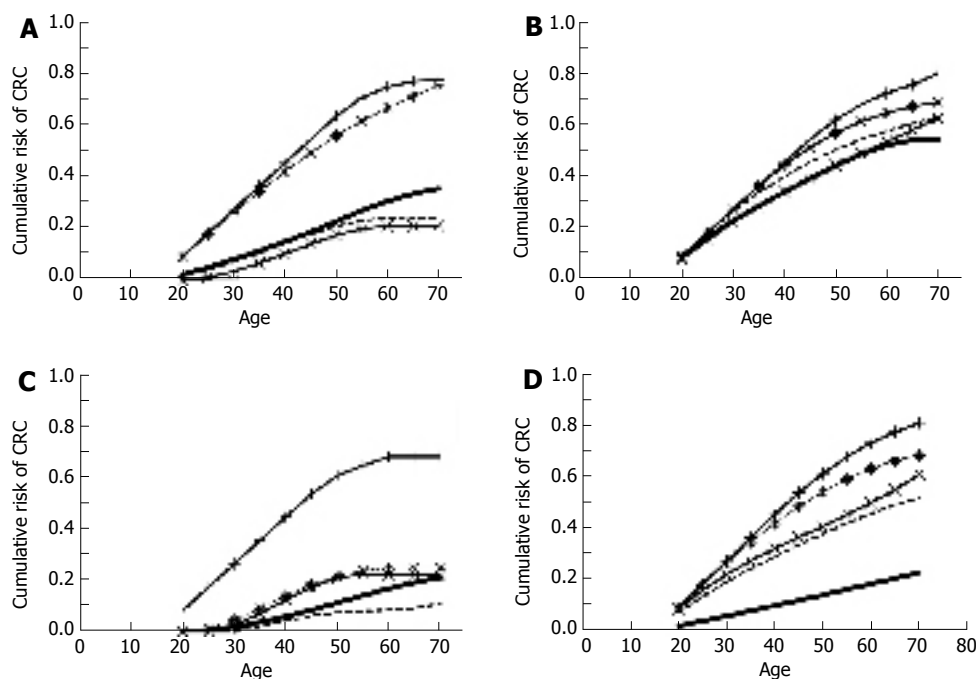
### Gender

Figure 4 demonstrates that males had a greater risk of developing CRC than females when compared to an equivalent age, gene mutation and smoking status. Males had a significantly higher risk of CRC than females (1.5-fold) for non-smoking subjects with the hMLH1

mutations (Figure 4A,  $P < 0.05$ ) and smokers (1.6-fold) (Figure 4B,  $P < 0.1$ ). Smoking males with the hMSH2 mutation also had a greater risk of CRC than hMSH2 mutation harboring females by 1.6-fold (Figure 4D,  $P < 0.01$ ), but the gender effect dissipated in nonsmokers (Figure 4C,  $P > 0.05$ ).

### Smoking dose response

A subset of the data which had an estimate of pack years was then modeled to determine if there was an association between lifetime quantities of cigarettes smoked and risk of CRC. The mean consumption was 24-pack years for patients who smoked cigarettes. Figure 5 demonstrates the



**Figure 5** Probability of developing CRC (cumulative lifetime risk) based on a model containing pack years for (A) females with hMLH1 mutation ( $P < 0.01$ ), (B) males with hMLH1 mutation ( $P < 0.05$ ), (C) females with hMLH1 mutation ( $P < 0.01$ ), and (D) males with hMSH2 mutation ( $P < 0.01$ ). The  $P$  values show the probability of failing to reject the null hypothesis that the data originate from identical populations. The number of pack years is as follows: - = 0, -x = 10, --- = 20, -\* = 30 and -+ = 40.

risk of developing CRC as a function of pack years. For female smokers with a hMLH1 mutation (Figure 5A), the CRC risk was only increased after 30-pack years ( $P < 0.01$ ), whereas for males (Figure 5B) the risk did not significantly increase until 40-pack years ( $P < 0.05$ ). Similarly, females with a hMSH2 (Figure 5C) only demonstrated a smoking effect at 40-pack years ( $P < 0.01$ ). Conversely, males with a hMSH2 mutation (Figure 5D) had an increased risk of developing CRC in a dose-dependent fashion in response to the number of cigarettes smoked over their lifetime ( $P < 0.01$ ).

## DISCUSSION

We have demonstrated herein that by using a fuzzy modeling approach, we could quantitatively predict the effect of environmental factors on risk of developing CRC in subjects who harbor a germline mutation for HNPCC. Importantly, we could calculate estimates for the impact of modifiable risk factors (i.e. smoking) on the occurrence of CRC in these high-risk patients and individualize the risk estimates by accounting for other major factors on the phenotypic variability in HNPCC patients: the mutated gene (hMLH1 versus hMSH2) and gender. Thus, we believe that these results may be a useful tool in patient counseling by providing concrete estimates of the impact of risk factor modification.

Our observations regarding the gene-environment joint effects were made possible by the remarkable resource represented by the Creighton Hereditary Cancer Center Registry. Although it is one of the oldest and largest HNPCC registries in the world, a conventional statistical approach to this dataset is not powerful enough to detect the gene-environment joint effect and the dose-response of smoking and CRC<sup>[15]</sup>, because the relatively small subgroup size markedly reduces statistical power in conventional (e.g. Cox proportional hazard modeling) statistical approaches. One approach to mitigate these

concerns is to increase the size of groups (e.g. to evaluate effect of age by increasing 40-59, 60-79, *etc.*). However, such large groups have clear disadvantages. For instance, a 41-year old individual and a 58-year old individual may be quite biologically/ clinically different and yet are in the same stratum. Fuzzy modeling enables partial membership functions. For instance, a 43-year old individual may be considered to be 80% in the 40-50 group and 20% belonging to the 30-39 category, whereas a 59-year old individual may be 65% in the 50-59 group and 35% within the 60-70 group. Thus, fuzzy modeling allows us to account for the heterogeneity, i.e., “shades of gray” that is a hallmark of clinical medicine.

There are several lines of evidence that support the biological validity of our findings with fuzzy modeling. Cigarette use is an important risk factor for CRC, and 12% of all CRC deaths are attributed to smoking<sup>[16]</sup>. Many studies indicate that cigarette smoking can increase the incidence of colon cancers by approximately two fold, however there are numerous contradictory reports<sup>[17-19]</sup>. These discordant data have been clarified by the demonstration that cigarette smoking may selectively increase the risk for DNA mismatch repair<sup>[20]</sup>. This may be related to the observation that smoking preferentially promotes microsatellite unstable (MSI-high) tumors. The molecular pathway is also seen in Lynch syndrome tumors<sup>[19]</sup>. For instance, Yang *et al.*<sup>[21]</sup> have recently reported that cigarette smoking increases the risk of developing MMR-deficient tumors by 3.1-fold. Additionally, Slattery *et al.*<sup>[22]</sup> demonstrated that smoking 20 cigarettes per day increases the risk of MMR-deficient tumors by 1.6-fold in men (95% CI = 1.0-2.5) and 2.2-fold in women (95% CI = 1.4-3.5). Furthermore, they have documented a dose-dependent relationship between smoking and colorectal cancer<sup>[22]</sup>. This dose-dependence underscores the plausibility of the cigarette-induced CRC risk. While we were able to discern an effect of smoking in our previous

study in HNPCC- cigarette smoking data set with Cox proportional hazard modeling, the lack of dose response raises concerns about the validity of the findings<sup>[15]</sup>. Using fuzzy modeling we demonstrated a much clearer relationship. For instance, our dose response model predicts that female smokers with a hMLH1 mutation who have at least 30-pack years of smoking will have an increase in the lifetime risk of CRC by 2.2- fold. A 3.3-fold increase is seen for females with an hMSH2 mutation after 40 pack years. Males with a hMSH2 mutation have a more linear increase in their lifetime risk of developing CRC as a function of pack years.

This ability to quantitate an individual's risk is of paramount clinical importance due to the variability in CRC presentation that is characteristic of HNPCC. For instance, some members of a kindred may develop CRC at age 25 and 65 while other members may never develop it. Given this heterogeneity, "one size fits all" approach to management (the current state of the art) is clearly inadequate. Indeed, previous attempts to determine the optimal cancer prevention strategy (prophylactic colectomy versus colonoscopic surveillance) have failed to conclusively demonstrate the superiority of any single approach<sup>[7]</sup>. Even determining the best colonoscopic intervals is unclear. While our group recommends annual colonoscopy starting at age 25<sup>[23]</sup>, a large number of negative examinations are expensive and have potential complications and may lead to patient complacency. Increasing surveillance intervals is fraught with danger given both the rapid adenoma to carcinoma transition and also the flat nature of the lesions, leading to a higher possibility of lesions being missed on colonoscopy<sup>[23]</sup>. The consequences of inadequate screening are underscored by the report of Jarvinen and colleagues<sup>[24]</sup>, who noted that over a 15-year observation period, 8.4% of HNPCC patients who did not undergo screening would die of CRC whereas none of those who were in a screening program can succumb to this malignancy. Indeed, in mutation positive subjects, development of CRC occur in 42% of the non-screened but only 18% in patients receiving screening ( $P < 0.02$ )<sup>[24]</sup>. Thus, implementation and adherence of a screening regimen are critical in protecting these high-risk patients against CRC.

It needs to be emphasized that the ability of fuzzy modeling to quantitate risk may be of considerable importance in counseling patients. For instance, being able to tell patients that their risk of CRC more than doubles with smoking may be more tangible than stating that smoking is detrimental to ones health, thereby providing a greater impetus for behavior modification. By accurately delineating risk, patients will be able to concretely identify modifiable risk factors and ascertain the impact of their lifestyle changes, thus providing positive reinforcement. In this regard, Halpert and colleagues<sup>[25]</sup> noted that genetic testing of HNPCC patients may have a profound effect upon motivation for cancer prevention strategies such as colonoscopy. Improved adherence with CRC screening regimens from genetic testing and counseling is also documented by Hadley and associates<sup>[26]</sup>. The malleability of CRC prevention behaviors in HNPCC patients is further highlighted by Adams and colleagues<sup>[27]</sup> who

documented the effect of socio-economic considerations on age of resection of CRC in HNPCC patients. Thus, we believe that added information obtained by fuzzy modeling may have a dramatic effect on patient behavior and thus outcomes. There are previous demonstrations of the efficacy of fuzzy logic to cancer risk stratification. Fuzzy logic has been used with impressive success to improve the sensitivity of tumor markers in diagnosing cancer<sup>[11-13]</sup>.

There are several limitations of this report that need to be acknowledged. As any modeling, the accuracy of the results is dependent on quality of the data inputted. Many of our patients did not have data to quantitate pack-years. Bias in the database due to patient/family report is possible (e.g. having cancer may influence recollection of tobacco use history). Since tobacco use is not a "standard" risk-factor for CRC, we do not think this will impact the results. With smoking, there is always concern about confounding from "competing causes of mortality"<sup>[28]</sup>. However we have recently shown that this effect is negligible for the smoking-CRC effect<sup>[29]</sup>. Finally, while our modeling accurately reflects our database, the algorithms need to be validated in other databases.

In conclusion, a fuzzy modeling approach represents a promising means of predicting the phenotypic heterogeneity in colorectal cancer presentation in HNPCC mutation carriers. The methodology may be an important tool in unraveling the gene-environment joint effects in hereditary cancer syndromes. Furthermore, this may serve as the basis for future paradigms that determine individualized cancer prevention strategies in subjects harboring an inherited risk.

## ACKNOWLEDGMENTS

The authors would like to acknowledge the statistical support of Dr. David Marx, Department of Statistics, University of Nebraska, Lincoln.

## REFERENCES

- 1 Jemal A, Murray T, Samuels A, Ghafoor A, Ward E, Thun MJ. Cancer statistics, 2003. *CA Cancer J Clin* 2003; **53**: 5-26
- 2 Grady WM. Genetic testing for high-risk colon cancer patients. *Gastroenterology* 2003; **124**: 1574-1594
- 3 Rubinstein WS, Roy HK. Practicing medicine at the front lines of the genomic revolution. *Arch Intern Med* 2005; **165**: 1815-1817
- 4 Kong S, Amos CI, Luthra R, Lynch PM, Levin B, Frazier ML. Effects of cyclin D1 polymorphism on age of onset of hereditary nonpolyposis colorectal cancer. *Cancer Res* 2000; **60**: 249-252
- 5 Roy HK, Lynch HT. Diagnosing Lynch syndrome: is the answer in the mouth? *Gut* 2003; **52**: 1665-1667
- 6 Ashwathnarayan R, Watson P, Lynch HT, Roy HK. Gene-environment interactions in hereditary nonpolyposis colorectal cancer: potentiation of colon cancer risk by tobacco use and hmlh-1 mutations. *Am J Gastroenterol* 2003; **98**: S112
- 7 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 Intern Med* 1998; **129**: 787-796
- 8 Dunlop MG, Farrington SM, Carothers AD, Wyllie AH, Sharp L, Burn J, Liu B, Kinzler KW, Vogelstein B. Cancer risk associated with germline DNA mismatch repair gene mutations. *Hum Mol Genet* 1997; **6**: 105-110

- 9 **Sanchez E.** Fuzzy logic and inflammatory protein variations. *Clin Chim Acta* 1998; **270**: 31-42
- 10 **Pannier AK, Brand RM, Jones DD.** Fuzzy modeling of skin permeability coefficients. *Pharm Res* 2003; **20**: 143-148
- 11 **Keller T, Bitterlich N, Hilfenhaus S, Bigl H, Loser T, Leonhardt P.** Tumour markers in the diagnosis of bronchial carcinoma: new options using fuzzy logic-based tumour marker profiles. *J Cancer Res Clin Oncol* 1998; **124**: 565-574
- 12 **Schneider J, Bitterlich N, Velcovsky HG, Morr H, Katz N, Eigenbrodt E.** Fuzzy logic-based tumor-marker profiles improved sensitivity in the diagnosis of lung cancer. *Int J Clin Oncol* 2002; **7**: 145-151
- 13 **Halm U, Rohde N, Klapdor R, Reith HB, Thiede A, Etzrodt G, Mossner J, Keller T.** Improved sensitivity of fuzzy logic based tumor marker profiles for diagnosis of pancreatic carcinoma versus benign pancreatic disease. *Anticancer Res* 2000; **20**: 4957-4960
- 14 **Steel RDG, Torrie JH.** Principles and Procedures of Statistics, a Biometrical Approach. 2nd ed. New York: McGraw-Hill, 1980
- 15 **Watson P, Ashwathnarayan R, Lynch HT, Roy HK.** Tobacco use and increased colorectal cancer risk in patients with hereditary nonpolyposis colorectal cancer (Lynch syndrome). *Arch Intern Med* 2004; **164**: 2429-2431
- 16 **Chao A, Thun MJ, Jacobs EJ, Henley SJ, Rodriguez C, Calle EE.** Cigarette smoking and colorectal cancer mortality in the cancer prevention study II. *J Natl Cancer Inst* 2000; **92**: 1888-1896
- 17 **Newcomb PA, Storer BE, Marcus PM.** Cigarette smoking in relation to risk of large bowel cancer in women. *Cancer Res* 1995; **55**: 4906-4909
- 18 **Heineman EF, Zahm SH, McLaughlin JK, Vaught JB.** Increased risk of colorectal cancer among smokers: results of a 26-year follow-up of US veterans and a review. *Int J Cancer* 1994; **59**: 728-738
- 19 **Neugut AI, Terry MB.** Cigarette smoking and microsatellite instability: causal pathway or marker-defined subset of colon tumors? *J Natl Cancer Inst* 2000; **92**: 1791-1793
- 20 **Slattery ML, Levin TR, Ma K, Goldgar D, Holubkov R, Edwards S.** Family history and colorectal cancer: predictors of risk. *Cancer Causes Control* 2003; **14**: 879-887
- 21 **Yang P, Cunningham JM, Halling KC, Lesnick TG, Burgart LJ, Wiegert EM, Christensen ER, Lindor NM, Katzmann JA, Thibodeau SN.** Higher risk of mismatch repair-deficient colorectal cancer in alpha(1)-antitrypsin deficiency carriers and cigarette smokers. *Mol Genet Metab* 2000; **71**: 639-645
- 22 **Slattery ML, Curtin K, Anderson K, Ma KN, Ballard L, Edwards S, Schaffer D, Potter J, Leppert M, Samowitz WS.** Associations between cigarette smoking, lifestyle factors, and microsatellite instability in colon tumors. *J Natl Cancer Inst* 2000; **92**: 1831-1836
- 23 **Lynch HT, de la Chapelle A.** Genetic susceptibility to non-polyposis colorectal cancer. *J Med Genet* 1999; **36**: 801-818
- 24 **Jarvinen HJ, Aarnio M, Mustonen H, Aktan-Collan K, Aaltonen LA, Peltomaki P, De La Chapelle A, Mecklin JP.** Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer. *Gastroenterology* 2000; **118**: 829-834
- 25 **Halbert CH, Lynch H, Lynch J, Main D, Kucharski S, Rustgi AK, Lerman C.** Colon cancer screening practices following genetic testing for hereditary nonpolyposis colon cancer (HNPCC) mutations. *Arch Intern Med* 2004; **164**: 1881-1887
- 26 **Hadley DW, Jenkins JF, Dimond E, de Carvalho M, Kirsch I, Palmer CG.** Colon cancer screening practices after genetic counseling and testing for hereditary nonpolyposis colorectal cancer. *J Clin Oncol* 2004; **22**: 39-44
- 27 **Adams J, White M, Barker G, Mathers J, Burn J.** Are there socio-economic inequalities in age of resection of colorectal cancer in people with HNPCC? *Fam Cancer* 2003; **2**: 169-173
- 28 **Chiang CL.** Competing risks in mortality analysis. *Annu Rev Public Health* 1991; **12**: 281-307
- 29 **Zisman AL, Nickolov A, Brand RE, Gorchow A, Roy HK.** Associations between the age at diagnosis and location of colorectal cancer and the use of alcohol and tobacco: implications for screening. *Arch Intern Med* 2006; **166**: 629-634

S- Editor Wang J L- Editor Wang XL E- Editor Bi L