

How well does family history predict who will get colorectal cancer? Implications for cancer screening and counseling

David P. Taylor, PhD¹, Gregory J. Stoddard, MPH², Randall W. Burt, MD^{3,4}, Marc S. Williams, MD^{1,5}, Joyce A. Mitchell, PhD¹, Peter J. Haug, MD^{1,6}, and Lisa A. Cannon-Albright, PhD^{7,8}

Purpose: Using a large, retrospective cohort from the Utah Population Database, we assess how well family history predicts who will acquire colorectal cancer during a 20-year period. **Methods:** Individuals were selected between ages 35 and 80 with no prior record of colorectal cancer diagnosis, as of the year 1985. Numbers of colorectal cancer-affected relatives and diagnosis ages were collected. Familial relative risk and absolute risk estimates were calculated. Colorectal cancer diagnoses in the cohort were counted between years 1986 and 2005. Cox regression and Harrell's C were used to measure the discriminatory power of resulting models. **Results:** A total of 431,153 individuals were included with 5,334 colorectal cancer diagnoses. Familial relative risk ranged from 0.83 to 12.39 and 20-year absolute risk from 0.002 to 0.21. With familial relative risk as the only predictor, Harrell's C = 0.53 and with age only, Harrell's C = 0.66. Familial relative risk combined with age produced a Harrell's C = 0.67. **Conclusion:** Family history by itself is not a strong predictor of exactly who will acquire colorectal cancer within 20 years. However, stratification of risk using absolute risk probabilities may be more helpful in focusing screening on individuals who are more likely to develop the disease. *Genet Med* 2011;13(5): 385–391.

Key Words: colorectal cancer, family history, risk prediction, genetic epidemiology, Utah Population Database

Colorectal cancer (CRC) is the second leading cause of death among cancers in the United States. In 2009, it was estimated that 147,000 cases would be newly diagnosed and that 50,000 deaths would be caused by the disease.¹ Because CRC often develops from precancerous polyps that can be identified and removed, it is one of the few cancers that can be prevented through appropriate screening.^{2,3} It has been estimated that more than half of deaths from CRC could be prevented through early detection.⁴ Increased surveillance in those at elevated risk

may lead to the detection of more cases and, therefore, a potentially greater mortality reduction than general surveillance of the population.⁵ However, based on known risk factors, predicting who will develop CRC is still a challenge.

Family history has often been cited as an important risk factor for CRC based on evidence that those with a positive family history for CRC have elevated risk compared with those with no family history of the disease,^{6–9} and evidence that a stronger family history results in even higher risks.¹⁰ The most commonly used measure for family history of CRC is “≥1 affected first-degree relative,” and in a recent, large, population-based study, the associated familial relative risk was estimated as 2.05 (95% confidence interval: 1.96–2.14).¹⁰ Current screening guidelines are informed by these types of familial relative risk studies and typically recommend that individuals with a positive family history be screened earlier and more frequently than those without.^{2,11,12} In addition, clinicians tend to rely on familial relative risk estimates rather than other types of risk representations such as absolute risk, even though absolute risk may be more easily interpretable and understood.^{9,13,14}

Although it has been established that a positive family history is associated with increased risk, an important question is whether increased familial relative risk is actually a clinically significant predictor of who will develop CRC. There are few published large prospective studies that assess familial risk and subsequent CRC diagnosis. The studies that do exist rely on self-reported family history and are limited to first-degree relatives.^{15,16} The primary purpose of this study is to assess how well family history predicts who will get CRC over a period of 20 years, using a large, retrospective cohort from the Utah Population Database (UPDB). Such information is critical for health policy organizations that address screening strategies and similarly important for practitioners who recommend screening, and genetic counselors who advise persons of cancer risk.

MATERIALS AND METHODS

The UPDB is a population-based, electronic genealogical resource that contains multiple linked data sources including statewide cancer registry records.¹⁷ It was created in the early 1970s with data from the Utah Family History Library and contains genealogies for the original Utah pioneers and their modern day descendants.^{18,19} While the original Utah Genealogy included records for 1.6 million persons,¹⁷ today the UPDB includes information for approximately 7 million persons, with some pedigrees >11 generations deep, although not all persons have linked genealogical data. The UPDB also includes a link to the Utah Cancer Registry (UCR), a statewide cancer registry established in 1966, which since 1973 has been part of the Surveillance, Epidemiology, and End Results (SEER) network of National Cancer Institute registries. Among those with cancer in the UCR, 94% link to ≥1 records in the UPDB, and 64.2%

From the ¹Department of Biomedical Informatics, University of Utah School of Medicine; ²Division of Epidemiology, Department of Internal Medicine, University of Utah School of Medicine; ³Huntsman Cancer Institute; ⁴Department of Internal Medicine, University of Utah School of Medicine; ⁵Clinical Genetics Institute, Intermountain Healthcare; ⁶Homer Warner Center for Informatics Research, Intermountain Healthcare; ⁷Division of Genetic Epidemiology, Department of Internal Medicine, University of Utah School of Medicine; and ⁸George E. Wallen Department of Veterans Affairs Medical Center, Salt Lake City, Utah.

David P. Taylor, MS, Department of Biomedical Informatics, University of Utah School of Medicine, Health Sciences Education Bldg, 26 South 2000 East, Suite 5700, Salt Lake City, UT 84112-5750. E-mail: d.p.taylor@utah.edu.

Disclosure: The authors declare no conflict of interest.

Submitted for publication September 13, 2010.

Accepted for publication November 9, 2010.

Published online ahead of print January 25, 2011.

DOI: 10.1097/GIM.0b013e3182064384

Table 1 Examples of colorectal cancer (CRC) family history constellations and corresponding familial relative risk estimates^a

| No. affected first-degree relatives | No. affected second-degree relatives | No. affected third-degree relatives | Familial relative risk (95% CI) |
|-------------------------------------|--------------------------------------|-------------------------------------|---------------------------------|
| 0 | 0 | 0 | 0.83 (0.81–0.86) |
| 0 | 0 | ≥3 | 1.08 (0.97–1.20) |
| 0 | 1 | 2 | 1.33 (1.13–1.55) |
| 1 | 0 | 0 | 1.76 (1.63–1.89) |
| 1 | 0 | ≥3 | 2.01 (1.61–2.47) |
| 1 | 1 | 0 | 1.88 (1.59–2.20) |
| 1 | 1 | ≥3 | 3.28 (2.44–4.31) |
| 2 | 0 | 0 | 2.96 (2.41–3.60) |
| 2 | 0 | ≥3 | 4.82 (3.18–7.02) |
| 2 | 1 | 1 | 1.80 (0.82–3.41) |
| 2 | 1 | ≥3 | 4.67 (2.72–7.47) |
| ≥3 | 0 | 0 | 2.96 (1.42–5.44) |
| ≥3 | 0 | 1 | 4.21 (1.82–8.30) |
| ≥3 | 0 | ≥3 | 9.63 (5.26–16.15) |
| ≥3 | 1 | 0 | 12.39 (7.08–20.12) |
| ≥1 (dx age <50 yr) | NA | NA | 3.31 (2.79–3.89) |
| ≥1 (dx age 50–59 yr) | NA | NA | 2.53 (2.24–2.85) |
| ≥1 (dx age 60–69 yr) | NA | NA | 2.22 (2.04–2.40) |

^aFull estimates are reported in Ref. 10.

CI, confidence interval.

have family information. Cancer records are coded by disease site according to the International Classification of Diseases of Oncology.²⁰ Information on site, stage, grade, age at diagnosis, histology, and patient survival are included. The UCR only reports independent primary cancers.

In contrast to many religious populations, individuals in the UPDB have been shown to be genetically representative of US white and northern European populations^{21–24} with a low-to-normal level of inbreeding.²⁵ Also, many are members of The Church of Jesus Christ of Latter-day Saints, which has religious proscriptions against the use of coffee, tea, alcohol, and tobacco. Consequently, much lower smoking rates may play a role in Utah being among states with the lowest rates of cancer.²⁶

This project used a subset of UPDB records representing a group of 2.3 million persons. These individuals were part of ≥3 generations of Utah genealogy data and descendants of original Utah pioneers. To protect the privacy of the study individuals, identifying information was not available to the authors. The Utah Resource for Genetic Epidemiology, created in 1982, governs access to the UPDB.²⁷ The Utah Resource for Genetic Epidemiology and the University of Utah Institutional Review Board granted approvals to conduct this research.

A retrospective cohort study design was used for this research. Considering that the latest cancer diagnosis information available in the dataset was from 2005, an observation period of 20 years was selected, with enrollment and family history assessment determined for the year 1985. The following were selection criteria: (1) individual's record was available in the

UPDB before 1986, (2) no record of death before 1986, (3) no record of CRC diagnosis before 1986, and (4) as of December 31, 1985, individual was between the ages of 35 and 80 years. Data were collected on the individual's age and family history as of December 31, 1985. Numbers of CRC-affected first-degree relatives, second-degree relatives, and third-degree relatives were gathered and numbers of CRC-affected first-degree relatives diagnosed between ages 50 and 69 years.

Previously, we reported familial relative risks for probands with various combinations, or constellations, of affected relatives with CRC, using the group of 2.3 million persons described earlier.¹⁰ Examples of constellations and their corresponding familial relative risks are listed in Table 1. A familial relative risk for each proband in this study as of the assessment date (December 31, 1985) was produced based on the proband's constellation of affected relatives in 1985.

During the observation period from 1986 to the end of 2005, data were collected on the years of occurrence of CRC diagnoses and deaths (from any cause) within the cohort. When an individual was diagnosed with CRC or died during the observation period, their record was censored during the year the event occurred. Cox regression was used to analyze the dataset, with CRC diagnosis as the dependent variable. The concordance statistic (or area under a receiver operating characteristic curve) is often used to assess the discriminatory power of a prediction model.^{28–31} Discriminatory power measures the ability of a model to distinguish between those individuals having a particular outcome and others without the outcome. It corresponds to the probability that a randomly selected individual who devel-

Table 2 Baseline characteristics and observed and expected CRC diagnoses for 20-yr observation period, 1985–2005

| | <i>n</i> (%) | No. w/CRC in observation period (%) | Expected no. CRC cases ^a |
|--------------------------------------|----------------|-------------------------------------|-------------------------------------|
| Total individuals | 431,153 | 5,334 (1.2) | |
| Age 35–49 yr | 163,277 (37.8) | 886 (0.5) | |
| Age 50–59 yr | 87,828 (20.4) | 1,249 (1.4) | |
| Age 60–69 yr | 104,420 (24.2) | 1,840 (1.8) | |
| Age 70–80 yr | 75,628 (17.5) | 1,359 (1.8) | |
| Familial relative risk | | | |
| 0 ≤ Familial relative risk < 1.0 | 402,317 (93.3) | 4,660 (1.2) | |
| 1.0 ≤ Familial relative risk < 2.0 | 19,299 (4.5) | 402 (2.1) | |
| 2.0 ≤ Familial relative risk < 3.0 | 8,238 (1.9) | 226 (2.7) | |
| 3.0 ≤ Familial relative risk < 4.0 | 1,250 (0.3) | 40 (3.2) | |
| 4.0 ≤ Familial relative risk < 9.0 | 41 (<0.1) | 4 (9.8) | |
| Familial relative risk ≥ 9.0 | 8 (<0.1) | 2 (25.0) | |
| Twenty-year absolute risk prediction | | | |
| 0 ≤ Absolute risk < 0.01 | 173,655 (40.3) | 990 (0.6) | 808 |
| 0.01 ≤ Absolute risk < 0.03 | 247,438 (57.4) | 3,991 (1.6) | 3,822 |
| 0.03 ≤ Absolute risk < 0.05 | 8,550 (2.0) | 286 (3.3) | 326 |
| 0.05 ≤ Absolute risk < 0.07 | 1,465 (0.3) | 61 (4.2) | 83 |
| 0.07 ≤ Absolute risk < 0.13 | 33 (<0.1) | 3 (9.1) | 3 |
| Absolute risk ≥ 0.13 | 12 (<0.1) | 3 (25.0) | 2 |

^aFrom absolute risk predictions based on 1981–1985 UPDB-specific incidence and mortality data.

ops the disease has a higher predicted risk than that of a randomly selected individual who does not develop the disease. The probability can range from 0.50, representing essentially a coin toss, to 1.00, representing perfect discrimination. A concordance statistic ≥ 0.70 is generally considered a threshold for a potentially useful model, but a value ≥ 0.80 may be a more reasonable level to provide adequate clinical utility.³²

An equivalent of the concordance statistic for use with Cox regression, Harrell's *C*,³³ was calculated for each model developed and compared. Because of the resource-intensive nature of the Harrell's *C* calculation, for models that included more than 100,000 individuals, Harrell's *C* was averaged across 10 random samples of 10% of the individuals in the model. Models were developed for familial relative risk as the sole predictor, age as the sole predictor, and familial relative risk and age included together. Familial relative risk and age were modeled as categorical variables.

As an alternative to using familial relative risk as the predictor in a Cox regression, absolute risk was also used. For each study individual, the absolute risk of developing CRC in the next 20 years was estimated using the individual's age and familial relative risk in 1985, according to the method by DuPont and Plummer.³⁴ This method also requires age-adjusted CRC morbidity rates and age-adjusted all-cause mortality rates to estimate the absolute risk. These rates were created directly from the UPDB population individually for the years 1981–1985 and then averaged. The purpose was to simulate risk estimates in 1985 as if it were a prospective study. Absolute risk was also modeled using categorical variables. The expected

numbers of individuals to develop CRC within the observation period among different levels of risk were estimated by summing the predicted absolute risk probabilities in each risk category.

Subgroup analyses were also performed by dividing the cohort into familial relative risk deciles, absolute risk deciles, and age groups. For each, Cox regression was performed, and Harrell's *C* was calculated for the highest risk decile, or age group, when compared with the lowest.

RESULTS

There were a total of 431,153 individuals included in the cohort. Baseline characteristics of these individuals are listed in Table 2, as well as numbers of CRC diagnoses. The range of familial relative risk was 0.83–12.39. The majority of individuals in this cohort (93.3%) had a familial relative risk <1.0. Less than 0.4% had a familial relative risk ≥ 3.0 . The range of 20-year absolute risk was 0.002–0.21, and the majority (57.4%) had a probability between 0.01 and 0.03. More than 2% had a 20-year absolute risk probability ≥ 0.03 . During the observation period, 5,334 individuals developed CRC. The age category (measured at baseline) with the most CRC diagnoses was 60–69 (1,840/5,334 = 34.5%). The percentages of observed CRC cases out of total individuals in each absolute risk category ranged from 0.6% (0 ≤ absolute risk < 0.01) to 25% (absolute risk ≥ 0.13).

Table 3 contains results of Cox regression and Harrell's *C* analyses. When familial relative risk was the only predictor,

Table 3 Results of Cox regression to predict diagnosis of colorectal cancer (CRC) based on (a) familial relative risk, (b) age, (c) familial relative risk and age, (d) absolute risk, (e) familial relative risk comparing the highest with lowest decile, and (f) absolute risk comparing the highest with lowest decile

| | Hazard ratio (95% CI) | Harrell's C |
|--------------------------------------------------------------|-----------------------|-------------|
| Familial relative risk | | |
| 0 ≤ Familial relative risk < 1.0 | Reference category | 0.53 |
| 1.0 ≤ Familial relative risk < 2.0 | 1.88 (1.70–2.08) | |
| 2.0 ≤ Familial relative risk < 3.0 | 2.50 (2.19–2.86) | |
| 3.0 ≤ Familial relative risk < 4.0 | 2.82 (2.06–3.85) | |
| 4.0 ≤ Familial relative risk < 9.0 | 10.65 (4.00–28.40) | |
| Familial relative risk ≥ 9.0 | 38.24 (9.56–152.94) | |
| Age (yr) | | |
| 35–49 | Reference category | 0.66 |
| 50–59 | 2.81 (2.58–3.06) | |
| 60–69 | 3.91 (3.61–4.24) | |
| 70–80 | 5.12 (4.70–5.57) | |
| Familial relative risk and age | | |
| 0 ≤ Familial relative risk < 1.0 | Reference category | 0.67 |
| 1.0 ≤ Familial relative risk < 2.0 | 1.67 (1.51–1.85) | |
| 2.0 ≤ Familial relative risk < 3.0 | 2.27 (1.98–2.59) | |
| 3.0 ≤ Familial relative risk < 4.0 | 2.79 (2.04–3.81) | |
| 4.0 ≤ Familial relative risk < 9.0 | 6.67 (2.50–17.77) | |
| Familial relative risk ≥ 9.0 | 28.43 (7.11–113.73) | |
| Age 35–49 yr | Reference category | |
| Age 50–59 yr | 2.76 (2.54–3.01) | |
| Age 60–69 yr | 3.84 (3.54–4.16) | |
| Age 70–80 yr | 4.99 (4.58–5.44) | |
| Absolute risk | | |
| 0 ≤ Absolute risk < 0.01 | Reference category | 0.64 |
| 0.01 ≤ Absolute risk < 0.03 | 3.48 (3.25–3.74) | |
| 0.03 ≤ Absolute risk < 0.05 | 7.64 (6.70–8.72) | |
| 0.05 ≤ Absolute risk < 0.07 | 9.52 (7.35–12.33) | |
| 0.07 ≤ Absolute risk < 0.13 | 21.85 (7.03–67.86) | |
| Absolute risk ≥ 0.13 | 80.08 (25.78–248.73) | |
| Familial relative risk (highest to lowest comparison) | | |
| Lowest decile (familial relative risk = 0.83) | Reference category | 0.54 |
| Highest decile (1.02 ≤ Familial relative risk ≤ 12.39) | 2.17 (2.00–2.35) | |
| Age (oldest to youngest comparison) | | |
| 35–49 yr | Reference category | 0.69 |
| 70–80 yr | 5.08 (4.66–5.53) | |
| Absolute risk (highest to lowest comparison) | | |
| Lowest decile (0.002 ≤ absolute risk ≤ 0.003) | Reference category | 0.78 |
| Highest decile (0.02 ≤ absolute risk ≤ 0.21) | 12.21 (10.47–14.24) | |

CI, confidence interval.

Harrell's $C = 0.53$. When age was the only predictor, the age group 35–49 years was used as the reference and Harrell's $C = 0.66$. The age group with the highest hazard ratio was 70–80 (5.12, 95% confidence interval: 4.70–5.57). Combining age and familial relative risk as predictors produced a Harrell's $C = 0.67$. Using absolute risk as the predictor, which is based on age, familial relative risk, and population-specific CRC incidence and all-cause mortality rates, produced a Harrell's $C = 0.64$.

In the subgroup analysis, when the highest decile of familial relative risk ($1.02 \leq$ familial relative risk ≤ 12.39) was compared with the lowest (familial relative risk = 0.83), Harrell's $C = 0.54$. For age, comparing those in the 70–80 years age group with those in the 35–49 years group, Harrell's $C = 0.69$. Harrell's $C = 0.78$ for the analysis comparing the highest absolute risk decile ($0.02 \leq$ absolute risk ≤ 0.21) with the lowest ($0.002 \leq$ absolute risk ≤ 0.003). Harrell's C statistic estimates the probability that, of two randomly chosen patients, the patient with the higher prognostic score will remain free of CRC longer than the patient with the lower prognostic score from the Cox regression model.³³ That is, the model result and the actual patient outcome were concordant, where the model correctly discriminated, 78% of the time in this Cox regression model.

DISCUSSION

We have described a retrospective cohort study that included 431,153 individuals aged 35–80 years at the beginning of the 20-year observation period. We are not aware of any other retrospective cohort or prospective studies of family history and CRC that have followed up this many individuals over this length of time. In addition, family histories of CRC were available electronically through a population-based electronic medical data resource as opposed to typically self-reported data.

Numerous studies have demonstrated increased familial relative risk for CRC in those with affected relatives.⁹ According to our analysis, however, family history as represented by a familial relative risk estimate is by itself not a good predictor (Harrell's $C = 0.53$) of exactly who will develop CRC in the next 20 years. Even when comparing the highest familial relative risk decile in the cohort ($1.02 \leq$ familial relative risk ≤ 12.39) with the lowest (familial relative risk = 0.83), Harrell's C was 0.54. When familial relative risk cutoffs were set even higher for the comparison, Harrell's C continued to decline, perhaps due to fewer numbers of cases. Familial relative risk is commonly used to communicate risk levels in the literature and among physicians and genetic counselors. However, with a disease such as CRC, it may not be commonly understood that although a familial relative risk estimate may be elevated, the corresponding absolute risk may still not be high. For example, hypothetically if a disease affects 10 of 1000 people with a particular risk factor and affects 1 of 1000 people without the risk factor, the relative risk is 10.0. Despite the seemingly large relative risk, the absolute risk for those with the risk factor is still only 10 of 1000 (1%).

In contrast to familial relative risk, age is a stronger predictor for CRC (Harrell's $C = 0.66$). Including familial relative risk in addition to age only improves the discriminatory power by 0.01, to 0.67. Absolute risk, which combines both familial relative risk and age, produced a Harrell's $C = 0.64$. Considering absolute risk uses the same variables and takes into account population-specific CRC incidence rates; it is not clear why this statistic was not higher. In the age subgroup analysis comparing those in the 70–80 years age group with those in the 35–49

years group, Harrell's $C = 0.69$. These findings illustrate that using familial relative risk in combination with age, or alternatively absolute risk, has moderate predictive value for CRC. However, in the absolute risk subgroup analysis where the highest decile was compared with the lowest, Harrell's C improved substantially to 0.78. Although one may question the clinical utility of this particular subanalysis, it is worth noting that the highest decile of absolute risk includes those with a 20-year risk of 0.02 and greater. For illustration, 0.02 is essentially the 20-year risk of a 50-year old with ≥ 1 CRC affected first-degree relative, so the highest decile of absolute risk includes more than just those at the extreme high end of risk based on family history.

To provide additional perspective on the levels of Harrell's C found in this study based on family history or family history in combination with age, a recent comprehensive risk prediction model for CRC that included a range of risk factors including family history produced a concordance statistic of 0.61.²⁸ This was based on validation in a population independent of the one used to build the model.

Despite the moderate Harrell's C value of models taking into account family history and age, the potential clinical value of a predictive model based solely on these risk factors is doubtful. However, it may be useful to consider aspects of the analyses presented in Tables 2 and 4 for decisions about appropriate screening. At the very highest levels of risk (familial relative risk ≥ 4.0 or absolute risk ≥ 0.07), relatively large percentages of individuals (e.g., 1 in 4 or 1 in 10) categorized by both familial relative risk and absolute risk end up developing CRC. However, at more moderate levels of risk, absolute risk tends to stratify individuals more appropriately. As an example, there were 4,660 CRC cases among 402,317 individuals with familial relative risk < 1.0 (1.2%). Absolute risk roughly divides the same number of individuals into two categories, absolute risk < 0.01 and $0.01 \leq$ absolute risk < 0.03 , where 3,991 cases were classified among 247,438 individuals (1.6%) in the latter category. In the former category, at the lowest level of absolute risk, there were 990 cases out of 173,655 individuals (0.6%). At higher levels of risk, there were 272 CRC cases that developed in 9,537 individuals with familial relative risk ≥ 2.0 (2.9%). There were 353 cases that developed in the 10,060 with absolute risk > 0.03 (3.5%). Absolute risk also has the benefit of facilitating the prediction of expected numbers of cases, and based on expected numbers of CRC cases in Table 2, it predicts fairly well how many individuals in each risk category are going to develop the disease in a 20-year time period.

This method of using absolute risk, based on age and family history, may be a reasonable way to quantify and stratify CRC risk, and these results demonstrate the possible utility. Particularly in the higher absolute risk categories, the numbers of cancers or precancers discovered through screening could potentially be much higher than in the general population. Considering the costs of screening, particularly colonoscopy, there are financial benefits in targeting and screening a smaller segment of the population and detecting a greater number of potential cases. It may also be easier to motivate those who are at increased risk to undergo screening, especially using more understandable absolute risk probabilities. In addition, the potential yield by percentage of appropriate screening increases as risk level increases.

Additional insights from absolute risk estimates may be gained from Table 4, which presents an adaptation of National Comprehensive Cancer Network (NCCN) screening guidelines for those with a positive family history of CRC.³⁵ Twenty-year absolute risk estimates are provided for 5-year age increments

Table 4 Twenty-year absolute risk estimates by age and history of CRC-affected first- and second-degree relatives (FDRs and SDRs)

| Patient age (yr) | "Average" risk | ≥1 affected FDR dx age 50–59 yr | ≥1 affected FDR dx <age 50 yr | ≥1 affected FDR dx ≥age 60 yr | Two affected FDRs dx any age | Two affected SDRs dx any age |
|-----------------------------------------|-----------------------------------|----------------------------------|---------------------------------------------------------------------|----------------------------------|---------------------------------------------------------------------|----------------------------------|
| 35 | 0.002 | 0.006 | 0.008 | 0.005 | 0.007 | 0.003 |
| 40 | 0.004 | 0.011 | 0.015 | 0.009 | 0.013 | 0.005 |
| 45 | 0.007 | 0.018 | 0.023 | 0.014 | 0.021 | 0.009 |
| 50 | 0.010 | 0.026 | 0.034 | 0.021 | 0.031 | 0.013 |
| 55 | 0.014 | 0.036 | 0.046 | 0.028 | 0.042 | 0.017 |
| 60 | 0.018 | 0.045 | 0.059 | 0.036 | 0.054 | 0.022 |
| 65 | 0.021 | 0.052 | 0.067 | 0.041 | 0.061 | 0.025 |
| 70 | 0.021 | 0.052 | 0.067 | 0.041 | 0.061 | 0.025 |
| 75 | 0.018 | 0.045 | 0.058 | 0.035 | 0.053 | 0.021 |
| 80 | 0.013 | 0.032 | 0.042 | 0.025 | 0.038 | 0.015 |
| Colonoscopy recommendation ^a | Every 10 yr starting at age 50 yr | Every 5 yr starting at age 40 yr | Every 3–5 yr starting at age 40 or 10 yr before the earliest CRC dx | Every 5 yr starting at age 50 yr | Every 3–5 yr starting at age 40 or 10 yr before the earliest CRC dx | Every 5 yr starting at age 50 yr |

^aAdapted with permission from The NCCN 3.2010 Colon Cancer Clinical Practice Guidelines in Oncology. National Comprehensive Cancer Network, 2010. Available at: <http://www.nccn.org>. Accessed September 3, 2010. Most recent and complete version of the guideline is available at www.nccn.org.

from 35 to 80 years according to notable family history patterns and for those at average risk (defined as having a familial relative risk of 1.0). NCCN colonoscopy recommendations for each family history category are noted as well. One may consider the absolute risk of a 50-year old with average familial relative risk as estimated from our dataset (20-year absolute risk = 0.01) as a reference point. Generally, there is consistency between the NCCN recommendations and the absolute risk patterns in that the most aggressive screening recommendations are associated with the highest levels of absolute risk. However, increased screening based on affected second-degree relatives may not be justified based on these data.

Although the absolute risk estimates in this research are based on familial relative risks that consider extended CRC-affected relatives (second- and third-degree relatives) in addition to affected first-degree relatives, our previous work has shown that the influence of extended relatives is relatively small; risk estimates are available that consider only first-degree relatives.¹⁰ Because patients often are not aware of the cancer history (or get it wrong) in their extended relatives, this may be important to consider.³⁶ In addition, considering the limited time clinicians have to obtain family history, not having to collect data on second-degree relatives would be beneficial to some degree. Further research could address the impact of limiting the familial relative risk estimates to only affected first-degree relatives in the absolute risk estimates and in the expected/observed numbers of cases by risk category.

The limitations of this study include the fact that we were unable to determine whether some individuals moved out of state during the 20-year observation period and, therefore, should have been censored for analysis; this limitation would have served only to lower our estimates of diagnosis rates from truly higher rates and does not change our conclusions. Similar to other UPDB-based studies,¹⁰ these results are generalizable to other populations of northern European origin but may not be

generalizable to populations with very different racial and ethnic compositions. There is a reliance on appropriate cancer coding, but the source of cancer data was a National Cancer Institute's Surveillance, Epidemiology, and End Results registry. Not all relatives of individuals may be represented in the UPDB genealogy, but we anticipate no bias in such representation. In addition, the incidence of CRC in Utah is the lowest in the United States.³⁷ Although CRC incidence rates are slightly different in men and women and also between sites (e.g., colon versus rectum), we did not distinguish by sex or by site in this study, consistent with the level of granularity of the previously generated familial relative risk estimates. It is also not known what screening may have occurred in the cohort, particularly in those at increased familial risk, and what effect this may have had in preventing CRC that would have otherwise occurred. This remains a possible minimal confounding factor based on observations in a yet to be published parallel study. Also, individuals with familial forms of CRC such as hereditary nonpolyposis CRC have not been excluded from this study because they may not be reliably identified and one may question whether this could skew the analysis. However, based on a previous study using individuals in the UPDB, only a small number met the Amsterdam I criteria (65 of 9458 cases or 0.7% of the cases), and none had a histology indicating familial adenomatous polyposis syndrome.³⁸ Despite the limitations identified, this study adds considerable definition and specifics as to how the relative risks, which to date have been used to establish screening strategies for those with a family history of this disease, actually play out over a 20-year period. These results should be carefully considered by health policy organization as they establish screening guidelines and by clinicians and genetic counselors as they deal with persons and families with familial colon cancer risk.

In conclusion, although previous studies have demonstrated increased relative risk among those with a family history of

CRC, this large retrospective cohort study has demonstrated that family history, without respect to age, is not a strong predictor of exactly which individuals will acquire CRC in the next 20 years, based on Cox regression and a measurement of concordance. It is important to keep in mind that even if a relative risk estimate may seem large, absolute risk may still be small if the incidence of a disease is low. When combined with age in an absolute risk estimate, family history does seem to improve concordance in a subgroup analysis to compare those at higher risk with those at very low risk. However, it is doubtful that a clinically useful statistical model for predicting who will acquire CRC at an individual level can be produced using just age and family history. Despite this, absolute risk predicts fairly well how many individuals in particular risk categories will develop the disease over a period of 20 years. Stratification of risk using absolute risk in a clinical setting could help target screening on those individuals who are more likely to develop the disease. Future work would include validating these absolute risk estimates in an independent population, performing a cost/benefit analysis to determine optimal screening recommendations based on risk levels, and providing a web-based tool for clinicians to estimate absolute risk based on a patient's current age and family history.

ACKNOWLEDGMENTS

The authors thank Steve Backus for database and software support and Jim Farnham for additional statistical guidance.

Research was supported by the Utah Cancer Registry, which is funded by contract N01-PC-35141 from the National Cancer Institute's Surveillance, Epidemiology, and End Results program with additional support from the Utah State Department of Health and the University of Utah. Partial support for all data sets within the Utah Population Database was provided by the University of Utah Huntsman Cancer Institute and the Huntsman Cancer Institute Cancer Center Support Grant, P30 CA42014 from the National Cancer Institute. Additional support was from R01, National Library of Medicine Grant LM009331 and the Huntsman Cancer Foundation (L.A.C.-A.), National Cancer Institute Grants R01-CA40641 and P01-CA73992 (R.W.B.), NIH National Center for Research Resources Grant 5UL1-RR025764 (G.J.S.), and an Intermountain Healthcare, Homer Warner Center for Informatics Research Fellowship (D.P.T.).

REFERENCES

- Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. *CA Cancer J Clin* 2009;59:225–249.
- Levin B, Lieberman DA, McFarland B, et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *Gastroenterology* 2008;134:1570–1595.
- Pignone M, Rich M, Teutsch SM, Berg AO, Lohr KN. Screening for colorectal cancer in adults at average risk: a summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med* 2002;137:132–141.
- Colditz GA, Atwood KA, Emmons K, et al. Harvard report on cancer prevention volume 4: Harvard Cancer Risk Index. Risk Index Working Group, Harvard Center for Cancer Prevention. *Cancer Causes Control* 2000;11:477–488.
- Hunt LM, Rooney PS, Hardcastle JD, Armitage NC. Endoscopic screening of relatives of patients with colorectal cancer. *Gut* 1998;42:71–75.
- Johns LE, Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. *Am J Gastroenterol* 2001;96:2992–3003.
- Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J Natl Cancer Inst* 1994;86:1600–1608.
- Hemminki K, Li X. Familial colorectal adenocarcinoma from the Swedish Family-Cancer Database. *Int J Cancer* 2001;94:743–748.
- Butterworth AS, Higgins JPT, Pharoah P. Relative and absolute risk of colorectal cancer for individuals with a family history: a meta-analysis. *Eur J Cancer* 2006;42:216–227.
- Taylor DP, Burt RW, Williams MS, Haug PJ, Cannon-Albright LA. Population-based family history-specific risks for colorectal cancer: a constellation approach. *Gastroenterology* 2010;138:877–885.
- Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2008;149:627–637.
- Rex DK, Johnson DA, Anderson JC, Schoenfeld PS, Burke CA, Inadomi JM. American College of Gastroenterology guidelines for colorectal cancer screening 2009 [corrected]. *Am J Gastroenterol* 2009;104:739–750.
- Gaismaier W, Gigerenzer G. Statistical illiteracy undermines informed shared decision making. *Z Evid Fortbild Qual Gesundheitswes* 2008;102:411–413.
- Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. *BMJ* 2003;327:741–744.
- Fuchs CS, Giovannucci EL, Colditz GA, Hunter DJ, Speizer FE, Willett WC. A prospective study of family history and the risk of colorectal cancer. *N Engl J Med* 1994;331:1669–1674.
- Murphy G, Shu XO, Gao YT, et al. Family cancer history affecting risk of colorectal cancer in a prospective cohort of Chinese women. *Cancer Causes Control* 2009;20:1517–1521.
- Cannon-Albright LA. Utah family-based analysis: past, present and future. *Hum Hered* 2008;65:209–220.
- Skolnick MH. Prospects for population oncogenetics. In: Mulvihill JJ, Miller RW, Fraumeni JF, editors. Genetics of human cancer. New York: Raven Press, 1977:19–25.
- Skolnick MH. The Utah genealogical data base: a resource for genetic epidemiology. In: Cairns J, Lyon JL, Skolnick MH, editors. Banbury report no 4; cancer incidence in defined populations. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1980:285–297.
- World Health Organization. International classification of diseases for oncology, 3rd ed. Geneva, Switzerland: World Health Organization, 2000.
- Cannon-Albright LA, Thomas A, Goldgar DE, et al. Familiality of cancer in Utah. *Cancer Res* 1994;54:2378–2385.
- McLellan T, Jorde LB, Skolnick MH. Genetic distances between the Utah Mormons and related populations. *Am J Hum Genet* 1984;36:836–857.
- Jorde LB, Shortleeve PA, Henry JW, Vanburen RT, Hutchinson LE, Rigley TM. Genetic analysis of the Utah population: a comparison of STR and VNTR loci. *Hum Biol* 2000;72:927–936.
- Cannon-Albright LA, Farnham JM, Thomas A, Camp NJ. Identification and study of Utah pseudo-isolate populations-prospects for gene identification. *Am J Med Genet A* 2005;137A:269–275.
- Jorde LB. Inbreeding in the Utah Mormons: an evaluation of estimates based on pedigrees, isonymy, and migration matrices. *Ann Hum Genet* 1989;53:339–355.
- Lyon JL, Gardner JW, Klauber MR, Smart CR. Low cancer incidence and mortality in Utah. *Cancer* 1977;39:2608–2618.
- Wylie JE, Mineau GP. Biomedical databases: protecting privacy and promoting research. *Trends Biotechnol* 2003;21:113–116.
- Park Y, Freedman AN, Gail MH, et al. Validation of a colorectal cancer risk prediction model among white patients age 50 years and older. *J Clin Oncol* 2009;27:694–698.
- Kurz DJ, Bernstein A, Hunt K, et al. Simple point-of-care risk stratification in acute coronary syndromes: the AMIS model. *Heart* 2009;95:662–668.
- Elmore JG, Fletcher SW. The risk of cancer risk prediction: "What is my risk of getting breast cancer"? *J Natl Cancer Inst* 2006;98:1673–1675.
- Freedman AN, Seminara D, Gail MH, et al. Cancer risk prediction models: a workshop on development, evaluation, and application. *J Natl Cancer Inst* 2005;97:715–723.
- Ohman EM, Granger CB, Harrington RA, Lee KL. Risk stratification and therapeutic decision making in acute coronary syndromes. *JAMA* 2000;284:876–878.
- Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA* 1982;247:2543–2546.
- Dupont WD, Plummer WD Jr. Understanding the relationship between relative and absolute risk. *Cancer* 1996;77:2193–2199.
- NCCN 3.2010 Colon Cancer Clinical Practice Guidelines in Oncology. National Comprehensive Cancer Network, 2010. Available at: <http://www.nccn.org>. Accessed September 3, 2010.
- 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;36:309–312.
- Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2008. *CA Cancer J Clin* 2008;58:71–96.
- Maul JS, Warner NR, Kuwada SK, Burt RW, Cannon-Albright LA. Extracolonic cancers associated with hereditary nonpolyposis colorectal cancer in the Utah Population Database. *Am J Gastroenterol* 2006;101:1591–1596.