The clinical validity of family history in risk classification of colorectal cancer

Leigh Jonah

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfilment of the requirements for the M.Sc. degree in Epidemiology

Department of Epidemiology and Community Medicine
Faculty of Medicine
University of Ottawa

ABSTRACT

Objective:

To determine the clinical validity of family history (FH) in colorectal cancer (CRC) risk classification.

Methods:

The Assessment of Risk of Colon Tumors In Canada case-control dataset was used to develop regression models associating risk factors with CRC in Ontario adults. Two regression models ('clinically-driven' based on a previously published tool, and data-driven) examined discrimination between CRC cases and controls, with and without the inclusion of FH as a risk variable. Discrimination was assessed using the area under the receiver operator characteristics curve.

Results:

For males, with the addition of FH, there were statistically significant yet quantitatively modest improvements in both models (3.7% clinically-driven, 6.8% data-driven). For females, while FH was a statistically significant predictor of CRC status in the data-driven model, the improvement in discrimination was not significant in either model.

Conclusion:

FH provides very small improvement in model discrimination beyond other standard CRC risk factors.

ACKNOWLEDGEMENTS

I would like to acknowledge and express my gratitude to the following people who made the completion of this project possible.

Dr. Brenda Wilson and Dr. Beth Potter, my supervisor and co-supervisor, for their knowledge, guidance, support, availability, and patience in completion of my thesis.

Dr. Zahra Montazeri for her assistance and advice with statistical analysis and her willingness to answer all my questions.

Dr. Julian Little for his expertise and assistance in solving dataset problems.

Dr. June Carroll for her expertise and insight.

Qendresa Hasanaj for her assistance and answers.

Silvia Visentin, Laure Tessier, and Christina Catley for their assistance and support, and for being such friendly co-workers.

The National Cancer Institute and the Colon Cancer Family Registry for the use of OFCCR data.

Lastly, to my family and friends for their encouragement, love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ARCTIC	Assessment of Risk of Colon Tumors in Canada
AUC	Area under the ROC curve
BMI	Body mass index
CCO	Cancer Care Ontario
CHD	Coronary heart disease
CI	Confidence interval
ColonCFR	Colon Cancer Family Registry
CRC	Colorectal cancer
CVD	Cardiovascular disease
FDR	First-degree relative
FH	Family history
FOBT	Fecal occult blood test
HNPCC	Hereditary non-polyposis colorectal cancer
HR	Hazard ratio
NCI	National Cancer Institute
NIH	National Institutes of Health
NSAID	Non-steroidal anti-inflammatory drug
OFCCR	Ontario Family Colon Cancer Registry
OR	Odds ratio
PY	Pack years
ROC	Receiver operator characteristics
RR	Relative risk
SDR	Second-degree relative

EXECUTIVE SUMMARY

Introduction

Having a family history (FH) of a disease is strongly associated with an individual's risk for many common complex diseases, such as colorectal cancer (CRC). For that reason, it is argued that because FH is associated with elevated risk of so many conditions, it could be a useful screening tool for general populations. While there is strong evidence available regarding disease risk associated with a positive FH and many cases of its inclusion in clinical practice guidelines, there is a lack of empirical evidence regarding its use in primary care and regarding its usefulness as a screening tool. For it to be used in such a manner, there is a need for evidence regarding its predictive accuracy.

Objectives

1) To assess the predictive accuracy of FH, measured through the improvements in risk classification gained with the inclusion of FH definitions into CRC risk models.
2) To identify an optimal performing definition of FH available from the dataset.

Methods

To address objective 1, two sets of risk classification models for males and females were developed using the Assessment of Risk of Colon Tumors In Canada (ARCTIC) case-control dataset. A clinically-structured risk model was developed based on an existing risk assessment model used by the National Cancer Institutes as the basis of their colorectal cancer risk tool. The second set of models was developed in an empirical data-driven manner, evaluating a broader range of risk factors available in the dataset. The improvements in risk prediction with the inclusion of FH were measured through changes in the area under the receiver operator characteristics (ROC) curve.

To address the second objective, four definitions of FH were developed from the available data, and assessed through area under the ROC curve (AUC). The definition identified with optimal performance in terms of risk prediction was incorporated into the data-driven model as the definition of FH.

Results

For males, the addition of FH created statistically significant, yet quantitatively modest improvements in risk model discrimination in both the clinically-driven and data-driven models (3.7% and 6.8%, respectively). For females, while FH was a significant predictor of FH in the data-driven model, the modest improvements in model discrimination did not reach statistical significance.

Of the four developed definitions of FH, 'Having a FH of CRC in at least a second-degree relative' had the greatest model discrimination (AUC 0.773 (male), and 0.644 (female)).

Conclusion

While FH is strongly associated with CRC risk, its added discriminatory ability in risk modeling is modest, even if statistically significant in males. The true value of any gain in discrimination lies in its clinical relevance, which in this case would require investigation of clinical utility, using, for example, risk re-classification analyses and decision analytic modelling. Risk re-classification modeling would permit the quantification of the changes in false positives and negatives that occur with the inclusion of FH. It is these changes that determine the clinical importance of improved model discrimination.

CHAPTER I - INTRODUCTION

Personalized Medicine

The idea of personalized medicine is the tailoring of preventive and treatment interventions according to individual risk or probability of response.[1] In effect, this is a version of stratified medicine,[2] using risk algorithms, molecular diagnostics, and/or patients' clinical profiles.[3] In principle, more accurate stratification should lead to higher rates of intervention effectiveness, lower incidence of preventable disease, reduction in or avoidance of adverse effects, and more effective use of health care resources. Although personalized medicine is most closely associated with genetics and the 'omics' sciences, the goal of improving risk or response stratification may be achievable through other biological markers and also family history.

Family History

Although family history (FH) is frequently conceived of as reflecting genetic risk, in fact it offers insight into a family's shared genetic, environmental, and cultural factors which may alter members' susceptibility or resilience to any number of medical conditions.[4] FH is known to be a risk factor for many chronic diseases such as cardiovascular disease, cancer, diabetes, and stroke (See Table 1).[6] In genetics terms, these are referred to as 'complex disorders', indicating a lack of simple inheritance patterns that suggest single gene etiologies.

There are multiple methods for defining FH, with varying levels of detail and complexity. Differing definitions could include the closeness of family (first-degree relative, second-degree relative), the number of affected relatives, age of diagnosis of affected family members, history of a specific cancer, such as CRC, or cancer in general, as well as other diseases that may be related.

Some have argued that, because FH is associated with elevated risk of so many conditions, it would be useful as a screening tool for general populations, to identify people at elevated risk who would benefit from targeted preventive interventions.[7] Following through on this idea, Yoon, Scheuner and colleagues[8] proposed a classification system to stratify risk (Figure 1), using FH, into three levels: high, moderate, and average (general population).

Table 1 - Summary of meta-analysis pooled relative risk estimates and 95% confidence intervals for different diseases and types of family history. [6]

Disease	N	At least 1 FDR	N	At least 2 FDR	N
Colorectal Cancer	58	2.24 (2.06, 2.43)	47	3.97 (2.60, 6.06)	10
Prostate Cancer	59	2.42 (2.25, 2.60)	50	4.27 (3.13, 5.84)	8
Breast Cancer	-	1.80 (1.70, 1.91)	52		-
Lung Cancer	29	1.83 (1.65, 2.03)	25	2.54 (1.78, 3.63)	7
Ovarian Cancer	33	2.85 (2.41, 3.37)	26	14.74 (5.78, 37.60)	3
Stroke	63	1.73 (1.52, 1.97)	36	1.69 (1.16, 2.46)	2
Multiple Sclerosis	32	14.63 (11.1, 19.4)	20	43.4 (5.4, 81.4)	1

FDR, first-degree relative
N= number of studies providing data on each subgroup.

This approach to classifying risk is attractive, but it requires valid definitions of the thresholds between risk strata. The authors went on to operationalize this classification system for a number of conditions,[8] but it has not been validated in general clinical populations. While FH has been represented in many risk prediction systems and clinical practice guidelines, the underpinning evidence base appears to be sparse. One tendency is for estimates of relative risk (RR) to be used to guide risk stratification. This (and similar metrics such as the odds ratio (OR) and hazard ratio (HR)) measures strength of association,[9] but does not provide an estimate of individual probability of disease. For the latter, FH needs to be approached as a 'test', and predictive accuracy metrics (sensitivity, specificity, predictive values, etc.) used to formally evaluate its performance.[9]

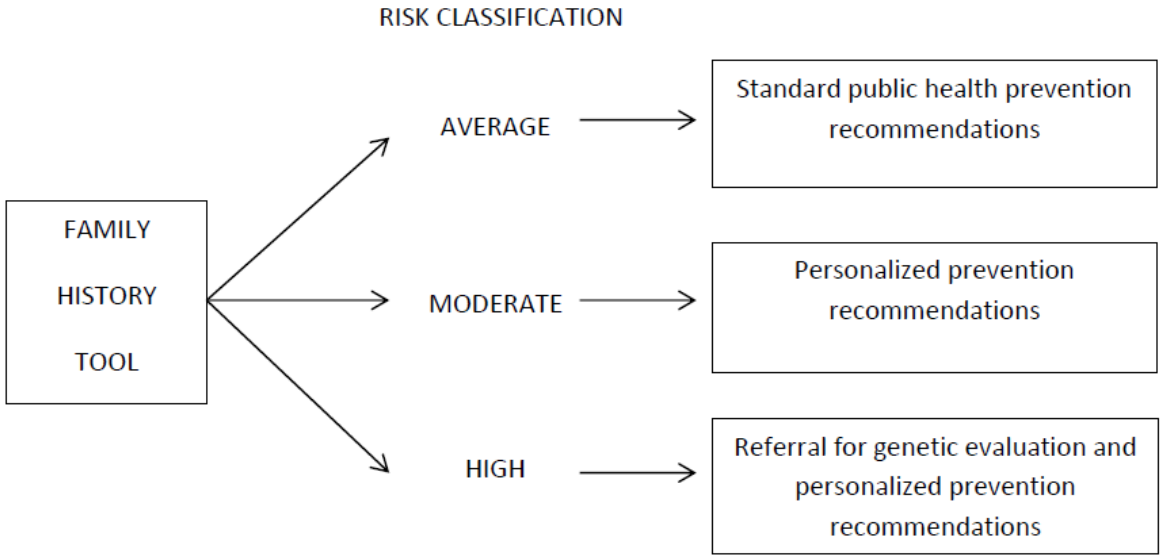


Figure 1 - Proposed scheme for using FH to guide and inform prevention activities[8]

In addition, for (genetically) complex disorders, FH by itself will carry only weak predictive information for individual disease risk:[10] if FH alone were strongly predictive of the risk of a disease, then the condition would not be 'complex', rather pedigree analysis would follow the pattern for highly penetrant single gene etiology (sometimes referred to as 'Mendelian'). Thus, for complex conditions, it is necessary to consider the 'added value' that FH information provides for disease prediction beyond standard risk factor assessment.[4] There remains little published information regarding the 'added value' of FH to standard risk factor assessment for many complex diseases.[11]

In 2009, the National Institutes of Health (NIH) held a State-of-the-Science panel on Family History and Improving Health and published recommendations[12] regarding further research relating to FH (Table 2). The NIH panel also concluded that for FH to be used as an evidence-based tool, its performance in terms of predictive ability and prognostic value needs evaluation.[12]

Table 2 - Research Recommendations – 2009 NIH State of the Science Panel [12]
1. Structure or characteristics of FH
What is a parsimonious series of questions (key elements) for use as a FH screening tool in primary care practice?
What are the environmental and lifestyle elements of an FH that are most useful in helping patients make positive changes in health-related behaviours?
What are the best methods and key elements to collect FH across several common disease entities?
How does the accuracy and completeness of FH information vary according to the setting in which it is collected (e.g. specialty care, primary care, community outreach, or the Internet)?
What is the optimal frequency for ascertaining and updating FH?
What are the best tools and methods for FH collection and interpretation?
What personnel and information technology resources and settings facilitate the collection of FHs that meet individual, community, and clinical goals?
What are the best statistical approaches to ascertain the benefit of 1 key element of FH relative to another element?
How does the definition of family in diverse racial, ethnic, religious, social, cultural, and economic population groups influence the collection and use of FH?
Do key elements of FH vary by race, ethnicity, religious belief, life stage, socioeconomic status, and culture?
How do family dynamics and health disorders affect and individual's awareness and ability to report on FH?
2. Process of acquiring a FH
Who is the best family informant to convey a FH?
To what extent do demographic factors modify an informant's ability to provide an accurate FH?
How might individuals, their families, and communities be best engaged in the collection of FH over time?
What are methods to minimize the time for collecting FH? Are there approaches to the assessment of FH across several office visits, self-administered questionnaires, ancillary personnel, or record linkage that are effective?
How do the clinician's knowledge, attitudes, beliefs, training, and skills influence the ability to collect, interpret, and use FH?
How might FH, including environmental and behavioural risk factors, be improved by a systematic, technology-supported approach?
What are optimal ways to use FH in a primary care setting to identify persons who can benefit from enhanced surveillance or referral to genetic services?
What are the key facilitators, incentives, and barriers for clinicians, individuals, families, and organizations for the collection of FH in primary care practice?
3. Expected outcomes of a FH interpretation
Besides disease risk assessment, what are the additional potential benefits in taking a thorough FH?
How and why does FH information change the behaviour of the clinician?
How are FH interpretation and findings best communicated to the individual/family to change health and disease prevention and detection behaviours over time? What strategies minimize potential harms?
What are the short- and long-term effects of inaccurate, misinterpreted, or unavailable FH information?
Can FH information be linked to genomic information or to important intermediate markers of common chronic diseases to predict change in outcome?
What are the short- and long-term effects on family dynamics of systematic FH taking in diverse populations and cultural settings?

"ACCE Framework"

The evaluative approach to personalized medicine applications, including FH, fits well within what is termed the ACCE framework – analytic validity, clinical validity, clinical utility, and ethical, legal and social implications (Table 3).[13] This was designed to clarify the evidence required in genetic test evaluation.

Table 3 - Elements and key components of evaluation framework for family history as a screening tool.[13]		
Elements	**Definition**	**Components**
Analytic validity	An indicator of how well a test or tool measures the property or characteristic (disease status among relatives) that it is intended to measure.	Analytical sensitivity Analytical specificity
Clinical validity	A measurement of the accuracy with which a test or tool identifies or predicts a clinical condition.	Clinical sensitivity Clinical specificity
Clinical utility	Degree to which benefits are provided by positive and negative test results (presence and absence of FH for disease).	Availability of effective interventions Health risks and benefits Economic assessment
Ethical, legal, and social implications	Issues affecting data collection and interpretation that might negatively impact individuals, families, and society.	Stigmatization Discrimination Psychological harm Risks to privacy and confidentiality

Analytic validity refers to whether a test measures what it is supposed to measure, in this case the accuracy of FH information actually captured by a FH tool. The metrics of interest are analytical sensitivity, how well the FH tool identifies family members that do have the disease, and analytical specificity, how well the FH tool identifies family members without the disease.[7]

Clinical validity refers to how well a test actually differentiates between individuals and groups with or without the attribute of interest, e.g. presence or absence of disease. In the current context, this means how accurately FH information stratifies patients according to actual disease risk. The key elements of clinical validity are clinical sensitivity and specificity, as well as positive and negative predictive values.[7] Ideally, these should be estimated from prospective studies in representative populations.

Clinical utility is the assessment of the likely usefulness of a clinically valid test in patient management – the degree to which health outcomes and/or resource use are improved when the test is integrated into practice. In the current context, this would examine whether using a FH tool under defined clinical circumstances leads to a higher proportion of people (those at higher risk of disease) actually receiving effective clinical or preventive interventions than they would otherwise have done

(fewer false negatives), and/or a lower proportion of people (those at lower risk) avoiding unnecessary interventions (fewer false positives). Full evaluation of clinical utility of FH information requires approaches such as randomized controlled trials comparing FH based stratification versus current practice (following through with altered clinical recommendations and interventions), or decision analysis to model the impact. Ideally, clinical utility evaluation should include examining risks of harm from a test (e.g. adverse psychological impact) and formal economic evaluation.

The final component of the ACCE framework is the consideration of ethical, legal and social implications which encompass broader issues surrounding the use of the tool or test. This commonly includes exploring issues such as stigmatisation, labelling, and discrimination, identifying ethical issues specific to the use of the test, clarifying legal and regulatory issues which may need to be addressed, and so forth.

This thesis focuses on applying the clinical validity component of the ACCE framework to FH information in prediction of risk of colorectal cancer (CRC).

Colorectal Cancer

CRC is the third most common cancer in Canada, with an estimated 23,300 new cases in 2012, and 9,200 deaths.[14] CRC is a complex disease where both environmental and genetic factors play an etiological role, with environmental factors and lifestyle being dominant.[15] It has been suggested that 70% of CRC deaths in the United States are potentially avoidable through modification of diet alone.[16] Of the evaluated risk factors for CRC, those supported by evidence are age (50 years or greater), being overweight or obese, diet (low in fiber, high in fats and red meat consumption), personal medical history (such as inflammatory bowel disease, diabetes, or previous history of colorectal polyps), use of aspirin and other non-steroidal anti-inflammatory drugs (NSAID), estrogen status and use of hormone replacement therapy, smoking, excess alcohol use and limited physical activity.[16–19]

Multiple genes have also been found to be associated with CRC, but in most cases they are not sufficient for CRC development, rather they interact with other genetic and environmental factors to increase or decrease CRC risk, or susceptibility to other etiological agents. The known genetic syndromes such as Lynch syndrome (previously known as hereditary non-polyposis colorectal cancer, HNPCC) and familial adenomatous polyposis (FAP) account for only about 5% of CRC cases, and these present with highly recognizable inheritance patterns. However, although most CRC cases are considered 'sporadic', there is undoubtedly familial clustering of the condition, the RR of even one affected first degree relative being estimated at 2.24 (95% Confidence Interval: 2.06-2.43).[6]

A review of longitudinal studies suggested that the sensitivity of FH alone in CRC risk prediction is around 13-14%, with a specificity of 92%.[9] There appear to be no evaluations of the clinical utility of using FH for CRC risk prediction.

Colorectal Cancer Risk Assessment

Deaths resulting from CRC are believed to be largely preventable. With regular screening, detection can occur at an initial stage, when symptoms are not present, but when treatment is most effective.[20,21] CRC screening can be thought of as a multi-stage process: age being the first phase, and fecal occult blood testing (FOBT) the second. FH fits into this process as an indicator of high risk, which might prompt alterations in the timing or nature of screening. In Ontario, the CRC screening guidelines, developed through Cancer Care Ontario (CCO), use age as the primary criterion for screening. In Ontario's Colon Cancer Check, the CRC screening program (Figure 2), individuals of 50 years of age or older are eligible to be screened through FOBT every 2 years.[22] Those at increased risk, defined as having one or more family members with CRC, begin screening through colonoscopy (rather than FOBT) at age 50, or 10 years earlier than the family member's age of diagnosis.[22]

The probability of survival is strongly associated with stage of detection - if caught early, there is a 90% chance of cure, whereas these chances decrease when detected at an advanced stage.[20] A systematic review of randomized control trials (RCTs) of screening using FOBT indicated a reduction in the relative risk of mortality of up to 25%.[23] Most screening programs are based on regular FOBT screening, applied after a threshold age (50 years in Ontario). This screening threshold shifts to an earlier point when the person is identified as high risk, through the presence of a FH of CRC. While population-based screening programs are evidence-based and effective in reducing mortality due to CRC, no program is perfect: some CRC cases are diagnosed clinically at a younger age than the screening thresholds, and some cases arise in the interval between screenings among those who participate in the screening program. Recently, there has been concern regarding the reduced effectiveness of colonoscopy screening in decreasing incidence and mortality of right-sided CRC;[24,25] FH may have the potential to act as a 'safety net' system that might catch these cases missed by colonoscopy.

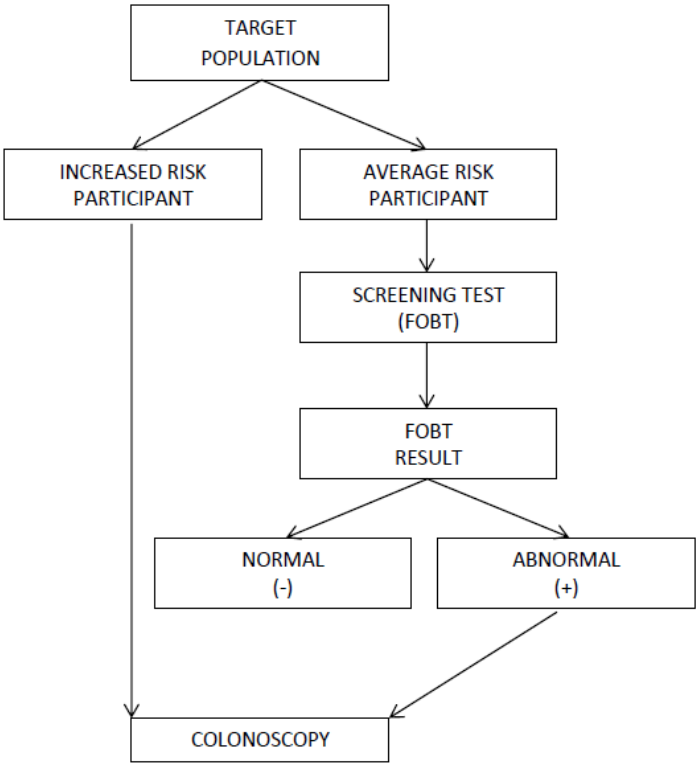


Figure 2 - Overview of CCO's Colon Cancer Check program.[22]
FOBT, fecal occult blood test.

Colorectal Cancer Risk Assessment Tools

Risk assessment tools, including tools that rely upon or incorporate FH, are often used to predict individual risk for common complex diseases within broad categories (risk strata), in order to tailor prevention strategies or guidance about screening.[7] For some diseases, there is a well-established standard risk assessment tool or system. For example, in cardiovascular disease (CVD), approaches derived from the Framingham cardiovascular risk profile are standard in North America.[26] The Framingham profile is used in clinical practice to determine an individual's risk of cardiac events, using the defined risk factors of age, sex, total cholesterol, high density lipoprotein (HDL) cholesterol, systolic blood pressure, blood pressure medication use, smoking, diabetes, and history of cardiovascular

events.[26] For CRC, there is no established equivalent. While there are knowledge syntheses that summarize evidence regarding risk factors consistently associated with CRC[16,27,28] no one model or set of predictors has become a standard for use in clinical practice.

A candidate model for CRC risk classification has recently been developed and is used as the basis for the National Cancer Institute's colorectal cancer risk evaluation tool.[29] Specifically, Freedman and colleagues developed colorectal cancer risk models for white men and women over the age of 50, using data from two United States case-control studies, and age-specific CRC incidence rates from Surveillance, Epidemiology, and End Results (SEER) data.[30,31] Freedman's model was validated in another dataset; a similar modest performance was found in the replication dataset.[31] This validated model could be useful as framework for model development in a Canadian population.

Colorectal Cancer Risk Classification

A better understanding of the value that FH adds to CRC risk assessment, beyond the standard risk factors (personal history, behavioural, and environmental), could help to advise existing screening guidelines. Small improvements in risk assessment could translate into better targeting of preventive interventions (such as screening) to those most likely to benefit, with intensity of intervention targeted to those at higher risk – potentially leading to important gains in population health.[4]

Risk assessment tools for CRC, as for any disease, must achieve useful levels of clinical validity – i.e., must be able to correctly stratify a patient's risk level.[4] Thus, to determine the independent value of FH in risk assessment, careful examination is required of relevant test metrics such as sensitivity, specificity, and the area under the receiver operator characteristics (ROC) curve (AUC).[4] Any research on the health impact of FH in CRC risk stratification (clinical utility) requires first determining that including FH in risk models indeed adds meaningfully to risk stratification.

Summary of Rationale

FH information may make an important contribution to risk stratification for CRC, and therefore could be seen as a relatively low cost tool for personalizing risk advice. However, using FH in this way requires evidence of its clinical validity. Currently, there is considerable evidence to support FH as a risk factor for many complex disorders, but very few empirical analyses of its value in individual disease risk stratification. This project makes use of a large, established dataset to add to this knowledge base.

Objectives

The overall goal of this project is to determine the clinical validity of FH in colorectal cancer risk classification. The specific objectives are as follows:

1. Informed by established risk factors, excluding FH, to develop multivariate models of CRC risk in a large dataset.
2. To assess the incremental improvements in risk classification gained when FH information is incorporated in the model developed in objective 1.
3. To identify the specific definition of positive FH that produces the most gain in risk classification.

CHAPTER II – METHODS

The project involved three parts:

I. Identification/development of a clinically driven risk-prediction model using standard risk factors and comparison of model discrimination estimates with and without FH information.
II. Modeling of FH definitions to identify the definition with the best performance in terms of risk discrimination.
III. Development of a best-fit data-driven model of risk with and without FH information and comparison of model discrimination estimates.

DATASET

This study used the Assessment of Risk of Colon Tumors in Canada (ARCTIC) case-control dataset, collected through the Ontario Family Colon Cancer Registry (OFCCR).[32,33] This is a region-specific subset of the international Colorectal Cancer Family Registry (Colon CFR), a collaboration of six centres from North America and Australia united as a single resource to support the study of the etiology, prevention, and the clinical management of CRC.[32]

Access to OFCCR data is managed through the National Cancer Institute's (NCI's) Informatics Support Center, which provided permission for its use in this project (Appendix A). REB approval for the secondary analysis of this existing dataset was obtained from the Ottawa Hospital Research Ethics board (OHREB) (Appendix B).

PARTICIPANTS AND RECRUITMENT STRATEGY

OFCCR

The OFCCR used two case recruitment strategies (termed in the original study "population based" and "clinic based"). For the current project, data pertaining only to participants recruited through the population-based strategy were included.[32] Because of a particular interest in genetic studies, the OFCCR sampling strategy targeted recruitment of individuals from families which appeared at higher risk of CRC. Potentially eligible cases were individuals aged 20-74 with CRC diagnosed between 1997 and 2007, identified through pathology reports (Table 4).[32,33] Permission to contact each potentially eligible case was sought from the primary care practitioner.[33] Where consent was received, the participant was sent the Family History Questionnaire (Appendix C) to collect information about first-degree relatives (FDRs), including cancer history.[33] This information was used to develop a pedigree, which was used to classify the case risk into one of three groups established by the OFCCR (see Table 5). Reflecting the OFCCR interest in higher risk families, 100% of individuals from the high and intermediate risk categories were selected to participate further in the study, compared with a 25% random sample of those in the low risk 'sporadic' category. [33] These sampled cases completed the Personal History Questionnaire and those who did so comprised the study sample of cases for OFCCR.

The Personal History Questionnaire (Appendix D) collected additional information on personal medical and health history, exposure to multiple risk factors, and family history information extended to second-degree relatives (SDRs).[33]

Controls were randomly selected from the general population using telephone subscribers' lists and frequency-matched by age (5 year age groups) and sex with the OFCCR case distribution.[33] Controls completed both the Family History and Personal History questionnaires.

Table 4 – CFR eligibility specifications by centre [32]	
Population-based recruitment	
CFR site/ascertainment source	
Cancer Care Ontario/ Ontario Cancer Registry	• Diagnosed in Ontario with incident CRC between the ages of 20-74 years. AND; • 100% of cases in Amsterdam I and intermediate-risk families and their FDR, half-siblings (at risk side), and blood relatives with HNPCC cancer (at-risk side) were recruited. • 100% of cases in multiple-case families (proband + with 2+ FDR or SDR <36y with an HNPCC* or other cancer) and all FDRs, half-siblings (at risk side) and blood relatives with HNPCC (at-risk side) with their adult FDR were recruited. • 25% of cases in single-case families, and for 25% of these cases, all FDR, half-siblings, and blood relatives with HNPCC with their adult FDR were recruited • All Cases with FAP were excluded.

*HNPCC (Lynch syndrome) – includes colorectal, endometrial, stomach, small bowel, gastroesophageal, pancreatic, biliary duct, liver, ovary, kidney, ureter, brain, and lymphoma cancer.

Table 5 – Criteria used to classify CRC risk of cases in the OFCCR[33]
High familial risk/Lynch Syndrome (Amsterdam Criteria)[34]
1. At least three relatives with colorectal cancer, one a first-degree relative to the other two, **and** 2. At least two successive generations affected with colorectal cancer, **and** 3. Colorectal cancer diagnosed under 50 years of age in at least one affected member, **and** 4. No Familial Adenomatous Polyposis (FAP)
Intermediate familial risk/Other risk
1. Proband has two relatives with any of the HNPCC cancers *and* two of the three are first-degree relatives, **or** 2. Any family member with an HNPCC cancer[b] ≤35 years of age, **or** 3. Proband <50 and relative with colon cancer <50 (first- or second-degree relative only), **or** 4. Proband ≤35 years of age, **or** 5. Proband with multiple primary colon cancers, **or** 6. Proband with other primary HNPCC cancer(s)[a], **or** 7. Proband has multiple polyps, **or** 8. Peutz-Jeghers or hamartomatous polyp, **or** 9. Juvenile polyp, **or** 10. Inflammatory bowel disease, **or** 11. Unusual colorectal cancer histologies[b], **or** 12. Proband is Ashkenazi Jewish
Low/Sporadic risk
1. All other colorectal cancer cases (probands) not classified as high or intermediate risk
[a] Colorectal, endometrial, gastric, small bowel, gastroesophageal, liver, pancreas, biliary tract, ovarian, kidney, ureter, brain, lymphoma. [b] Carcinosarcoma, adenosquamous, spindle cell, metaplastic, choriocarcinoma, signet ring, undifferentiated, trophoblastic differentiation, small cell neuroendocrine carcinoma

ARCTIC

The ARCTIC dataset was comprised of Cases and Controls from the OFCCR, with the further inclusion criteria of completion of the Family History and the Personal History questionnaires.

The OFCCR sampling design allowed for complete recruitment of participants aged 20-69. However, in practice, recruitment of participants 70-74 did not reach target. Only 575 of an estimated 1175 (49%) were actually recruited (Appendix E). In response, statisticians at the OFCCR altered the sampling weight for this group. In order to avoid complications for the current study, participants aged 70-74 were excluded.

SAMPLE STRUCTURE

The original dataset comprised 9035 participants who had completed the epidemiological questionnaire. The analysis sample was reduced to limit the impact of unequal probabilities of selection based on participant recruitment in the OFCCR.[33] A strategy to identify appropriate cases for inclusion, and to help mitigate possible selection bias, was developed through discussion with the thesis advisory committee (TAC) and via communication with experienced researchers.

Participants were limited to those who were recruited in the population-based recruitment step for the OFCCR recruitment strategy, excluding those in the clinic based sampling, as clinic-based sampling aimed to recruit cases and families meeting Amsterdam criteria for Lynch syndrome.[34] The sample was then limited to only the first identified case, and not those subsequently identified through the initial case's family. Participants ≥70 years of age were removed from the sample, due to the incomplete recruitment of this age group (as above). Of the remaining potential participants, 12 were misclassified (as case with no CRC, and control with CRC) and were removed.

Participants sampled in Ontario encompass a broad range of racial backgrounds. Unlike the participant sample used by the Freedman model, our sample was not limited to Caucasians, or those of European descent.

Any participants who had Familial Adenomatous Polyposis (FAP) and those who met Amsterdam I/II criteria (Table 6) were excluded. Lynch syndrome, identified using the Amsterdam criteria, and FAP are known genetic syndromes that present as highly recognizable inheritance patterns. Participants with these syndromes were excluded from analysis as the analysis aimed to investigate CRC in the average risk population.

Table 6 - Amsterdam I/II Criteria for Lynch Syndrome[34]

Amsterdam I criteria
There should be at least 3 relatives with CRC and all following criteria should be present: o One should be an FDR of the other two o At least 2 successive generations affected o At least 1 diagnosis of CRC before age 50 o FAP cases excluded o Tumors should be verified through pathological examination
Amsterdam II criteria
There should be at least 3 relatives with a Lynch syndrome-associated cancer (CRC, cancer of the endometrium, small bowel, ureter, or renal pelvis) o One should be an FDR of the other 2 o At least 2 successive generations should be affected o At least 1 diagnosis before age 50 o FAP cases excluded o Tumors should be verified through pathological examination

FDR, first-degree relative; CRC, colorectal cancer; FAP, familial adenomatous polyposis.

VARIABLES

Main Outcome

The main outcome was colorectal cancer. The dataset included a dichotomous CRC status variable: affected and unaffected. Affected cases were ascertained and confirmed by the OFCCR through pathology reports and histology samples.

Independent Variables

As three sets of models were developed for the three sections of the analysis (Table 7), the independent variables are identified separately for each.

Table 7 – Analysis models

Name	Model structure
Clinical Model	Model based on an existing, clinically-structured CRC risk assessment model,[30] using available variables from the ARCTIC dataset.
Family History Model	Investigation of varying definitions of FH available in the dataset to assess which definition has superior performance in terms of risk discrimination.
Data-Driven Model	Risk model developed with only variables identified as significant predictors from a data-driven model selection.

1) Clinical Model

For this analysis, we used the existing National Cancer Institute CRC risk assessment tool (hereafter referred to as the "Freedman model")[30] to select variables for the regression analysis. The Freedman model comprised separate risk models with different variables for men and women. Existing variables in the ARCTIC dataset were operationalized and coded to match the Freedman definition as closely as possible (Tables 8a,b).

Table 8a – Male – Freedman and Clinical model				
Risk Factor	**Freedman Definition**	**Categories**	**Clinical model Definition**	**Categories**
Polyp and Screening History	Sigmoidoscopy and/or colonoscopy and polyp history	Sigmoidoscopy and/or colonoscopy in last 10 years, and no history of polyps. No sigmoidoscopy and/or colonoscopy in last 10 years. Sigmoidoscopy and/or colonoscopy in last 10 years and history of polyps. Sigmoidoscopy and/or colonoscopy and polyps unknown.	Not included	N/A
Physical Activity	current leisure-time activity, hours/week	0 > 0 and ≤ 2 > 2 and ≤ 4 4	current leisure-time activity, hours/week	0 (REF) > 0 and ≤ 2 > 2 and ≤ 4 4
NSAID use		Non-user Regular user*		Non-user (REF) Regular user**
Smoking	cigarettes/day	Never smoker > 0 and < 11 ≥ 11 and ≤ 20 > 20	Pack Years	Never Smoker (REF) >0 and <20 20-29 ≥ 30
Smoking	years of smoking	Never smoker > 0 and < 15 ≥ 15 and < 35 ≥ 35		
Vegetable Consumption	Servings/day	< 5 ≥ 5	Servings/day	< 5 (REF) ≥ 5
Body Mass Index	kg/m^2	≤ 24.9 (under/normal) 25.0 to ≤ 30 (overweight) ≥ 30 (obese)	kg/m^2	≤ 24.9 (REF) 25.0 to ≤ 30 ≥ 30
Family History	Number of FDRs with CRC	0 1 ≥ 2	Number of FDRs with CRC	0 (REF) 1 ≥ 2

(REF) Indicates reference group.
* at least 3x/week for at least 1 month in previous year
**2x/week for at least 1 month at any point in the past
NSAID, non-steroidal anti-inflammatory.

Table 8b – Female – Freedman and Clinical model				
Risk Factor	**Freedman Definition**	**Categories**	**Clinical model Definition**	**Categories**
Polyp and Screening History	Sigmoidoscopy and/or colonoscopy and polyp history	Sigmoidoscopy and/or colonoscopy in last 10 years, and no history of polyps. No sigmoidoscopy and/or colonoscopy in last 10 years. Sigmoidoscopy and/or colonoscopy in last 10 years and history of polyps. Sigmoidoscopy and/or colonoscopy and polyps unknown.	Not included	N/A
Physical Activity	current leisure-time activity, hours/week	0 > 0 and ≤ 2 > 2 and ≤ 4 4	current leisure-time activity, hours/week	0 (REF) > 0 and ≤ 2 > 2 and ≤ 4 4
NSAID use		Non-user Regular user*		Non-user (REF) Regular user**
Estrogen Status	Within last 2 years	Positive Negative	Current estrogen status	Positive Negative
Vegetable Consumption	Servings/day	< 5 ≥ 5	Servings/day	< 5 (REF) ≥ 5
Body Mass Index	kg/m^2	≤ 24.9 (under/normal) 25.0 to ≤ 30 (overweight) ≥ 30 (obese)	kg/m^2	≤ 24.9 (REF) 25.0 to ≤ 30 ≥ 30
BMI-Estrogen Interaction		N/A	Not Included	N/A
Family History	Number of FDRs with CRC	0 1 ≥ 2	Number of FDRs with CRC	0 (REF) 1 ≥ 2

(REF) Indicates reference group.

* at least 3x/week for at least 1 month in previous year

**2x/week for at least 1 month at any point in the past

NSAID, non-steroidal anti-inflammatory; BMI, body mass index.

Age

As cases and controls were age matched within 5-year age groups (by frequency matching), age, in five-year intervals, was included in all models to control for any effect of inexact matching during recruitment. The age variable was derived from self-reported date of birth in the Family History Questionnaire (Appendix C). The 5-year age groups were created using participant age (at time of questionnaire) and correspond to the age groups used for control selection. Participants under 40 were grouped in one category because of small numbers.

Since age was included in the model to control for the unequal age distributions in cases versus controls, and since sampling was carried out based on age, the results for this variable are not meaningful as a potential risk factor.

Body mass index (BMI)

This variable was derived from the self-reported height and weight question (in the Personal History Questionnaire (Appendix D)) using the formula BMI= weight (kg)/height^2 (m^2).[35]

Polyp and Screening History

This variable was excluded from the analysis. In the Freedman model, this variable indicated a history of colorectal polyp in the preceding 10 years, which necessarily included differentiation between respondents who had, and had not, had a screening test (fecal occult blood test (FOBT), sigmoidoscopy and/or colonoscopy). Although this variable could have been derived from items in the Personal History Questionnaire, we concluded that, in the Canadian context, it would have acted as a marker of cancer detection rather than as a risk factor. The screening practices between the United States, where the Polyp and Screening History variable was developed (in the Freedman model), differ to those in Ontario. In the United States colonoscopy is

used as a primary screening method, whereas in Ontario colonoscopy is a secondary step in CRC screening, following positive FOBT, and would be indicative of a higher probability of disease. This assumption was explored in a sensitivity analysis (see Results).

Physical activity

This variable was derived from items in the Personal History Questionnaire (Appendix D, page 109). It was not possible to match the Freedman approach to defining and classifying this variable directly. Instead, a new composite physical activity variable was created, based on self-reported activity in the respondent's most recent age decade (determined from participant's age at time of questionnaire completion). In doing so, we had to take account of a large amount of missing data.

To assign a summary value of hours/week of physical activity, we combined responses to participation in an activity (yes/no), with the weekly duration (Figure 3). These were summed across activities. A response of 'No' for participation in an activity was assigned a value of 0 hours/week. A response of 'Yes' for participation in an activity accompanied by missing data for duration was assigned as 'missing'.

While this process helped to minimize the amount of missing data, there were still a large proportion of participants with missing values for the 'nearest decade' (16.9%). For participants with missing values for the most recent age decade, we imputed the level of physical activity using the responses for the next most recent age decade, where this was available. With imputation in this manner, the proportion of participants with missing data was reduced to 8.6%.

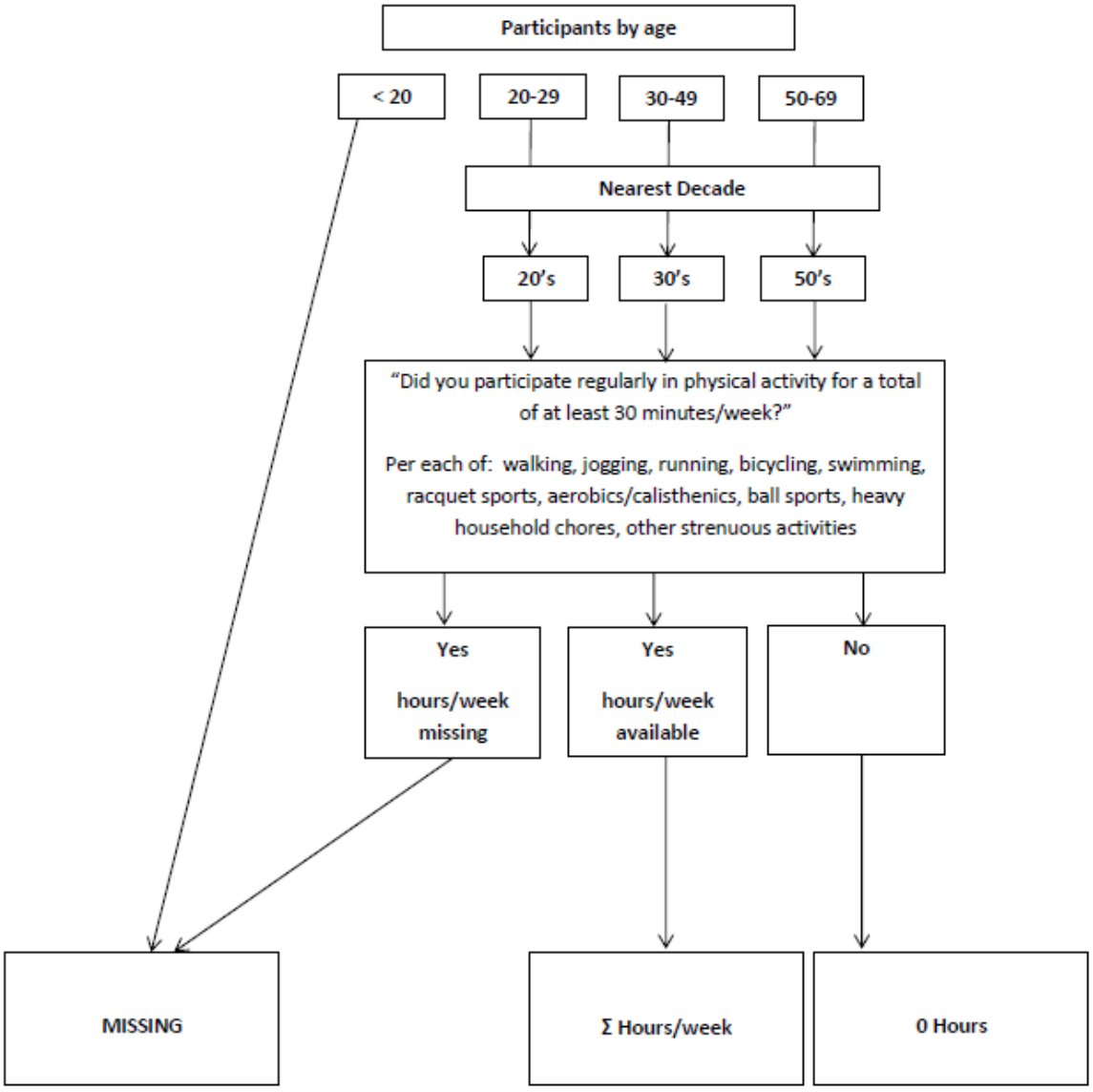


Figure 3 – Process for composition of physical activity variable (hours/week).

Non-steroidal anti-inflammatory medication (NSAID) use

This variable was derived from items in the Personal History Questionnaire. The Freedman model defined "regular use" as at least 3 times a week for at least one month in the previous year. [30] For the Clinical model, 'regular users' were defined as those who reported use of Aspirin and/or

Ibuprofen for at least twice a week for at least a month at any point in the past, all others being classed as 'non-regular users'.

Vegetable consumption

This variable was derived from items in the Personal History Questionnaire and was converted into servings/day to match the Freedman model.[30,36]

Smoking

This variable was derived from items in the Personal History Questionnaire. While the Freedman model contained two smoking variables, based on number of cigarettes smoked, and on years of smoking, these could not be separated in the ARCTIC dataset. A single variable of pack years was created which incorporated the length (years) and intensity (cigarettes smoked/day) components. Pack-years (PY) was calculated based on the definition from the National Cancer Institute[37]:

$$PY = \frac{(\#\ \text{cigarettes/day} \times \text{years smoked})}{20}$$

Thresholds for categories of pack-years were obtained from another study modeling CRC risk using Canadian data.[38] Smoking, in pack-years was divided into 4 categories: never smoker, >0-<20, 20-29, and ≥30.

Estrogen status

This variable (applicable to females only) was derived from an item in the Personal History Questionnaire. While use of hormone replacement therapy (HRT) was available, menstrual status was not available and could not be calculated. Therefore, based on the average age of menopause

in Canada, 51 years,[39,40] we assumed that respondents under 51 years were estrogen positive, and those 51 and over were estrogen negative, with the exception of those taking HRT, who were recorded as estrogen positive. We conducted a sensitivity analysis to test the effect of this assumption.

BMI-Estrogen interaction

A BMI-Estrogen interaction variable was planned, as it was included in the Freedman model, based on the BMI and estrogen variables defined above.

Family history

This variable was derived from items in the Family History Questionnaire. The coding used in the Freedman model was matched, i.e., number of FDRs with CRC.

2) Family History Model

This model was developed to examine the performance of different criteria for defining a person as having a positive FH of CRC, with the intention of using the definition with the highest AUC in the data-driven model to follow. Four possible definitions (Table 9) were created.

Table 9 - Different minimum criteria applied for classification as having 'positive' family history

Definition	Required degree of relationship	Required number of relatives with CRC
1	First	≥2
2	First or second	≥2
3	First	≥1
4	First or second	≥1

CRC, colorectal cancer.

These four definitions are somewhat nested, in that an individual meeting definition 1 (for example) would also meet definitions 2, 3, and 4. The regression models of the FH definitions were controlled for age, using the established 5-year age categories (<40, 40-44, 45-49, 50-54, 55-59, 60-64, and 65-69 years).

3) Data-driven model

The last set of models was developed in a data-driven manner. Risk factors for CRC were identified through literature review, and models (stratified by sex) best fitting the dataset were developed. The candidate variables (Table 10) were operationalized based on Freedman model definitions and other existing definitions of the risk factors. Variables included in the final model were selected, through backward elimination, as significant at $p<0.05$.

The following variables were examined, in addition to those already developed for the Clinical model. Each was selected because of evidence of association with CRC risk, on the basis of the literature review.

Diabetes

This variable was derived from an item in the Personal History Questionnaire. Diabetes has been indicated as a medical condition that can increase the likelihood of CRC.[27,41]

Inflammatory bowel diseases

This variable was derived from items in the Personal History Questionnaire, and was coded 'yes' if a respondent indicated having at least one of Crohn's disease or ulcerative colitis, conditions associated with an increased risk of CRC.[27,28]

Red meat consumption

This variable was derived from an item in the Personal History Questionnaire. Servings per week were calculated, and categorized into four classes similar to those used in analysis of CRC risk using Canadian data.[42]

Alcohol consumption

This variable was derived from items in the Personal History Questionnaire (Appendix D, page 115), using a process similar to that for creating the physical activity variable. A composite variable was created, based on self-reported consumption in the respondent's most recent age decade, accounting for the large amount of missing data.

The questionnaire collected information regarding consumption of beer, cider, wine, sake, sherry, port, spirits and liquor mixed drinks. To assign a summary value of drinks/week, responses of consumption (yes/no) and quantity (number of drinks consumed per week) were combined. If a participant had indicated 'No' for consumption of that alcohol type, a value of 0 drinks/week was added. If the participant indicated 'Yes' and the number of drinks was missing, then the consumption of that alcohol type was set to missing. A sum of drinks/week was created for each participant with valid data (i.e., no missing data on the variables making up the composite) – this sum was assumed to be representative of the participant's true alcohol consumption during the most recent decade. This method, while providing useful values of drinks/week for each participant, still resulted in a large amount of missing data.

To minimize the amount of missing data, imputation, in the same manner as for the physical activity variable, was used to estimate some participants' levels of alcohol consumption; i.e., we used values for previous decades when information was missing for the most recent

decade. With this imputation process, the proportion of participants with missing data was reduced from 21.3% to 6.2%.

Alcohol consumption, in drinks/week, was categorized to correspond with categories used in previous analysis of OFCCR data.[43]

Education

This variable was derived from an item in the Personal History Questionnaire, and was included as a marker of socio-economic status.

Family history

In contrast to the definition used in the Clinical model, the Data-driven model used as its FH variable the one determined to have the highest AUC in the FH model above.

Table 10 – Additional candidate variable definitions for data-driven analysis

Variable	Division
Diabetes	No Yes
Inflammatory Bowel Disease	No Yes
Red Meat Consumption	<1 servings/wk 1-3 servings/wk 4-6 servings/wk ≥7 servings/wk
Alcohol Consumption	0 drinks/wk 1-2 drinks/wk 3-7 drinks/wk >7 drinks/wk
Education	< High School High School Grad Some Post-Secondary Post-Secondary Grad
FH - Any FDRs or SDRs with CRC	No Yes

FH, family history; FDR, first-degree relative; SDR, second-degree relative; CRC, colorectal cancer

STATISTICAL ANALYSIS

SAS software (version 9.3) was used for all analyses.

Weighting

All analyses were weighted using the sampling weights derived by the primary OCCFR investigators, to account for the unequal probability of selection in the sampling design: for cases, sampling weights were either set as 1.0 or 4.0 to account for undersampling of lower risk cases; and for controls, all weights were 1.0 because there was no undersampling (Appendix E). Analysis weights were calculated from the sampling weights so that the sum of the sampling weights in each analysis was equal to the sample size:

$$\text{Analysis weight} = \frac{\text{Sampling weight}}{\text{Mean sampling weight}}$$

Univariate Analysis

Descriptive statistics of the dependent variable and the predictor variables considered for all analyses were calculated according CRC status, stratified by sex. Chi-square tests were used to assess the differences between cases and controls.

Multivariable Logistic Regression

Three sets of models were developed:

1) For the Freedman-based clinical model, base models without FH, for both sexes, were created. The enhanced clinical model then incorporated Freedman's definition of FH.

2) The second set of regression models (stratified by sex) was created to determine the performance of four definitions of FH, available from the dataset.
3) The last set of regression models (stratified by sex) was data-driven: we created an "ideal" model, to which FH was then incorporated. The ideal model was considered to be that with the best model discrimination.

Model performance (for the clinical and data-driven model sets) was assessed using 1) likelihood ratio tests and 2) comparison of areas under the ROC curve (AUC).

Sub-analysis and Sensitivity Analyses

Sub-analyses were planned to examine the performance of the base models and base+FH models when the population was stratified by age (<40 and ≥40).

Sensitivity analyses were planned to examine effects of excluding participants meeting definite or probable 'Amsterdam' criteria for Lynch syndrome (see Table 6), and for the effects of including the 'Polyp and Screening History' variable. We also planned a sensitivity analysis to examine the potential misclassification of estrogen status for women near the average age of menopause.

Risk re-classification analyses

In the original proposal, we planned to construct risk re-classification tables and calculate the Net Reclassification Index (NRI) to evaluate the effect of FH on risk stratification. However, the difficulty of this approach in case control data became apparent, because the reclassification analysis uses the regression to calculate each participant's probability of disease as a basis for comparison of different models. The probability of CRC in the ARCTIC dataset was set by the study

design (essentially 50%), and this is reflected in the regression equation. In all models, all participants have a higher estimated probability of disease than would be expected for a sample truly representative of the population. This impacts the "screening test" comparison of the models. While the NRI and risk re-classification tables can be calculated, this calculation in a case-control population would not be meaningful. Difficulties arise with the NRI when there is matching between cases and controls, such as the age and sex matching in our data – because of this matching, we would observe overweighting in NRI method.[44]

The sample for analyses included 3328 participants. Figure 4 summarizes the sample reduction process detailed in the methods.

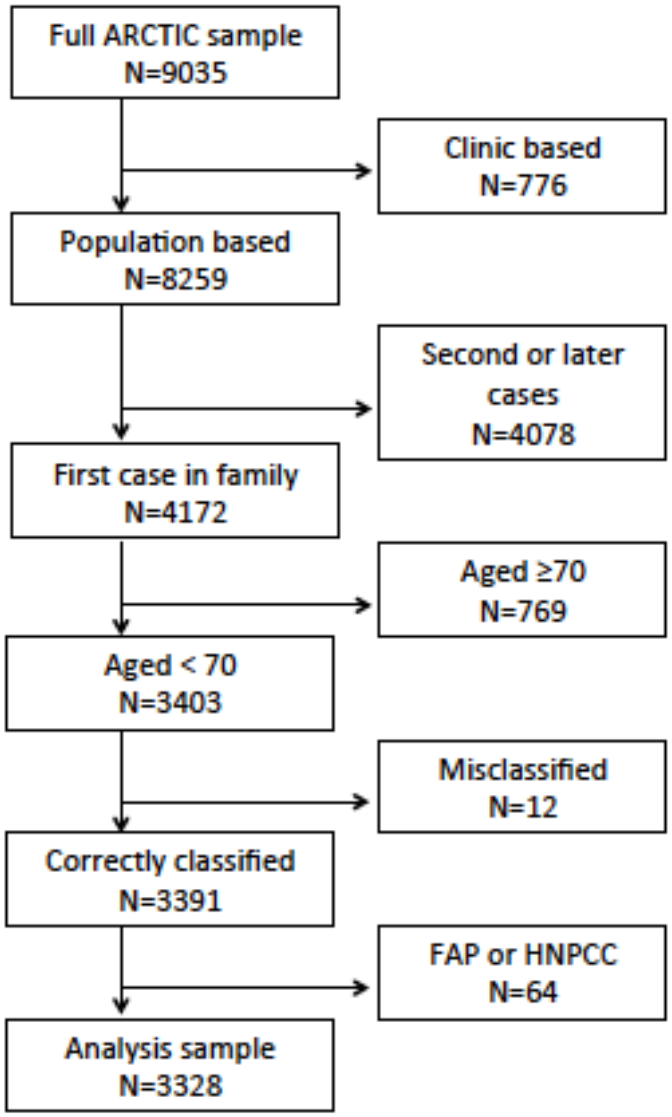


Figure 4 –Sample selection process

Analyses were stratified by sex (Figure 5). Missing observations were excluded from univariate analyses. The 'Final Sample' for regression analysis excluded participants with missing observations for any of the included model variables.

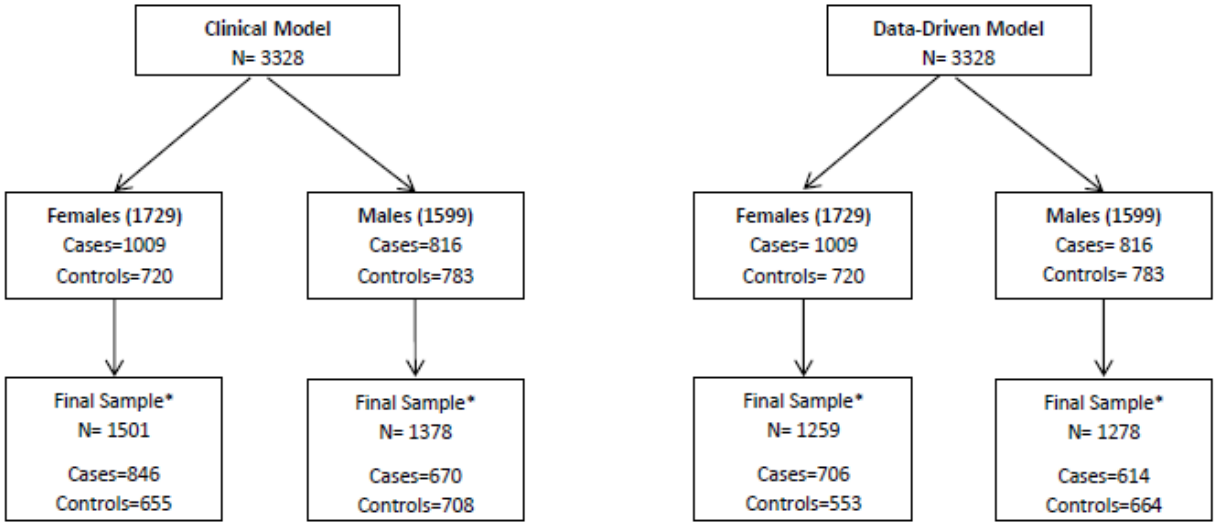


Figure 5 – Sample details for Clinical and Data-driven modeling.
*Final sample after excluding participants with missing data on included variables.

1) CLINICAL MODEL

Among males (Table 11a), cases compared to controls were more likely to be younger (30.1% vs. 10.1%) and to be overweight or obese (75.4% vs. 64.7%). Cases were more likely to have a FH of CRC, defined as one or two and greater affected FDRs, (20.9%) than controls (9.2%).

Table 11a - MALE – Comparison of Clinical model by cases and controls						
Variable	**Division**	**Cases (N=816)**		**Controls (N=783)**		**X^2 Test p-value**
		Un-weighted N (%)	**Weighted %***	**Un-weighted N (%)**	**Weighted %***	
BMI	Under/Normal	207 (25.91)	24.58	275 (35.35)	35.35	<0.0001
	Overweight	408 (51.06)	53.03	376 (48.33)	48.33	
	Obese	184 (23.03)	22.38	127 (16.32)	16.32	
Physical Activity	0 hrs/wk	64 (8.83)	9.07	74 (9.95)	9.95	0.6650
	>0 and ≤2 hrs/wk	97 (13.38)	13.81	102 (13.71)	13.71	
	>2 and ≤4 hrs/wk	78 (10.76)	10.48	91 (12.23)	12.23	
	>4 hrs/wk	486 (67.03)	66.64	477 (64.11)	64.11	
NSAID use	Yes	32 (3.93)	4.40	48 (6.13)	6.13	0.1264
	No	782 (96.07)	95.60	735 (93.87)	93.87	
Vegetables	<5 Servings/day	776 (97.98)	98.08	759 (98.70)	98.70	0.3620
	≥5 Servings/day	16 (2.02)	1.92	10 (1.30)	1.30	
Smoking	Never Smoker	312 (40.47)	37.21	280 (36.99)	36.99	0.9985
	>0 and <20 PY	223 (28.92)	29.61	226 (29.85)	29.85	
	20-29 PY	82 (10.64)	10.78	80 (10.57)	10.57	
	≥ 30 PY	154 (19.97)	22.40	171 (22.59)	22.59	
Age	<40	72 (8.82)	5.58	13 (1.66)	1.66	<0.0001
	40-44	100 (12.25)	7.89	30 (3.83)	3.83	
	45-49	200 (24.51)	16.67	36 (4.60)	4.60	
	50-54	84 (10.29)	11.16	113 (14.43)	14.43	
	55-59	90 (11.03)	13.84	170 (21.71)	21.71	
	60-64	118 (14.46)	20.39	197 (25.16)	25.16	
	65-69	152 (18.63)	24.48	224 (28.61)	28.61	
FH	0 FDRs with CRC	619 (75.86)	79.09	711 (90.80)	90.80	<0.0001
	1 FDRs with CRC	163 (19.98)	17.93	67 (8.56)	8.56	
	2+ FDRs with CRC	34 (4.17)	2.98	5 (0.64)	0.64	

*Weighted proportion to account for sampling design
Counts may not add up to total sample as missing data were excluded.
BMI, body mass index; NSAID, non-steroidal anti-inflammatory; PY, pack years; FH, family history; FDR, first-degree relative; CRC, colorectal cancer.

Among females (Table 11b), cases were more likely to be younger than controls (30.4% vs. 17.7% under 50 years), and less likely to use NSAID medication regularly (4.6% vs. 8.8%). There were no statistically significant differences between female cases and controls in BMI, levels of physical activity, vegetable consumption, estrogen status, or FH.

Table 11b - FEMALE – Comparison of Clinical model by cases and controls						
Variable	**Division**	**Cases (N=1009)**		**Controls (N=720)**		**X^2 Test p-value**
		Un-weighted N (%)	**Weighted %***	**Un-weighted N (%)**	**Weighted %***	
BMI	Under/Normal	493 (50.67)	49.59	367 (52.13)	52.13	0.6146
	Overweight	302 (31.04)	31.90	213 (30.26)	30.26	
	Obese	178 (18.29)	18.50	124 (17.61)	17.61	
Physical Activity	0 hrs/wk	81 (9.00)	9.59	77 (11.37)	11.37	0.5998
	>0 and ≤2 hrs/wk	112 (12.44)	13.27	89 (13.15)	13.15	
	>2 and ≤4 hrs/wk	146 (16.22)	15.27	111 (16.40)	16.40	
	>4 hrs/wk	561 (62.33)	61.87	400 (59.08)	59.08	
NSAID use	Yes	46 (4.58)	4.59	63 (8.75)	8.75	0.0006
	No	959 (95.42)	95.41	657 (91.25)	91.25	
Vegetables	<5 Servings/day	927 (95.17)	95.39	671 (93.72)	93.72	0.1385
	≥5 Servings/day	47 (4.83)	4.61	45 (6.28)	6.28	
Estrogen Status	Negative	294 (29.46)	33.21	255 (35.71)	35.71	0.3006
	Positive	704 (70.54)	66.79	459 (64.29)	64.29	
Age	<40	79 (7.83)	5.86	30 (4.17)	4.17	<0.0001
	40-44	98 (9.71)	8.51	49 (6.81)	6.81	
	45-49	194 (19.23)	16.01	48 (6.67)	6.67	
	50-54	119 (11.79)	12.37	114 (15.83)	15.83	
	55-59	124 (12.29)	12.72	125 (17.36)	17.36	
	60-64	175 (17.34)	19.16	174 (24.17)	24.17	
	65-69	220 (21.80)	25.38	180 (25.00)	25.00	
FH - Freedman	0 FDRs with CRC	799 (79.19)	82.20	620 (86.11)	86.11	0.1013
	1 FDRs with CRC	177 (17.54)	15.23	88 (12.22)	12.22	
	2+ FDRs with CRC	33 (3.27)	2.57	12 (1.67)	1.67	

*Weighted proportion to account for sampling design
Counts may not add up to total sample as missing data were excluded.
BMI, body mass index; NSAID, non-steroidal anti-inflammatory; FH, family history; FDR, first-degree relative; CRC, colorectal cancer.

REGRESSION MODELING

Age-adjusted ORs are presented, to account for age differences that were apparent from the sampling design, and multivariate-adjusted ORs are presented for each variable included in the model.

Males

The results of the multivariate model show that of the six covariates (+ FH) included in the base model, only three (BMI, NSAID use, and Age) and FH were significantly associated with CRC.

For the male population, Table 12a shows that the odds of having CRC were statistically significantly elevated for those who were 'overweight' (OR=1.89) or 'obese' (OR=2.27) compared to those of 'normal/underweight'. The odds were statistically significantly reduced for those regularly using NSAID medication (OR=0.55). Participants with a positive FH of CRC had significantly elevated odds of disease (OR= 2.42 for one affected FDR, 5.51 for ≥2 affected FDRs).

Table 12a – Male Clinical model – age and multivariate adjusted ORs

Variable	Age-Adjusted OR[a]		Multivariate-Adjusted OR[b]	
	OR	95% CI	OR	95% CI
BMI				
Under/Normal	1.00	-	1.00	-
Overweight	1.859	1.431, 2.415	1.894	1.450, 2.475
Obese	2.205	1.577, 3.082	2.266	1.608, 3.192
Physical Activity				
0 hrs/wk	1.00	-	1.00	-
>0 and ≤2 hrs/wk	1.171	0.727, 1.887	1.363	0.833, 2.229
>2 and ≤4 hrs/wk	0.913	0.555, 1.502	0.980	0.584, 1.644
>4 hrs/wk	1.080	0.726, 1.606	1.135	0.751, 1.713
NSAID Use				
No	1.00	-	1.00	-
Yes	0.598	0.361, 0.990	0.553	0.328, 0.931
Vegetables				
< 5 Servings/day	1.00	-	1.00	-
≥5 Servings/day	1.392	0.487, 3.977	1.548	0.519, 4.617
Smoking				
Never Smoker	1.00	-	1.00	-
>0 and <20 PYs	1.229	0.929, 1.625	1.218	0.914, 1.624
20-29 PYs	1.362	0.926, 2.002	1.378	0.925, 2.054
≥ 30 PYs	1.274	0.940, 1.727	1.226	0.894, 1.681
Age				
<40	1.00*	-	1.00	-
40-44	0.567	0.240, 1.344	0.577	0.240, 1.388
45-49	1.089	0.473, 2.506	1.015	0.435, 2.368
50-54	0.239	0.110, 0.518	0.207	0.094, 0.456
55-59	0.194	0.091, 0.413	0.154	0.071, 0.335
60-64	0.271	0.128, 0.570	0.226	0.105, 0.485
65-69	0.258	0.123, 0.543	0.219	0.102, 0.468
FH				
0 FDRs with CRC	1.00	-	1.00	-
1 FDRs with CRC	2.366	1.645, 3.404	2.415	1.663, 3.506
2+ FDRs with CRC	5.199	1.729, 15.631	5.508	1.814, 16.718

a - Adjusted for age, due to important differences in the age distributions of cases and controls that may reflect the sampling design

b - Adjusted for all covariates.

* Unadjusted results for age.

OR, odds ratio; CI, confidence interval; BMI, body mass index; NSAID, non-steroidal anti-inflammatory; PYs, pack years; FH, family history; FDR, first-degree relative; CRC, colorectal cancer.

Females

The results of the multivariate model show that of the six covariates (+ FH) included in the base model, only three (physical activity, NSAID use and Age) were significantly associated with CRC, while FH was not significant. Table 12b shows that the odds of having CRC were statistically significantly elevated for those women who engaged in the highest levels of physical activity (OR=1.46). The effect of the highest level of physical activity was noted in the multivariate adjusted analysis, but there was no effect in the age-adjusted analysis. The odds were significantly reduced for those who regularly used NSAID medication (OR=0.48). There was no significant effect of BMI, vegetable consumption, estrogen status, or FH.

Interaction term

The BMI*Estrogen interaction was tested in the clinical model for females. The interaction term was neither significant in the model that excluded FH (p=0.5172) nor in the FH-inclusive model (p=0.5223). The interaction was removed as it did not contribute significantly to the model.

Table 12b – Female Clinical model – age and multivariate adjusted ORs				
Variable	**Age-Adjusted OR[a]**		**Multivariate-Adjusted OR[b]**	
	OR	**95% CI**	**OR**	**95% CI**
BMI				
Under/Normal	1.00	-	1.00	-
Overweight	1.105	0.864, 1.413	1.121	0.873, 1.438
Obese	1.024	0.763, 1.375	1.057	0.783, 1.426
Physical Activity				
0 hrs/wk	1.00	-	1.00	-
>0 and ≤2 hrs/wk	1.350	0.863, 2.113	1.384	0.881, 2.175
>2 and ≤4 hrs/wk	1.309	0.854, 2.007	1.359	0.882, 2.094
>4 hrs/wk	1.401	0.976, 2.011	1.457	1.009, 2.104
NSAID Use				
No	1.00	-	1.00	-
Yes	0.482	0.311, 0.748	0.478	0.307, 0.745
Vegetables				
< 5 Servings/day	1.00	-	1.00	-
≥5 Servings/day	0.812	0.516, 1.278	0.812	0.511, 1.288
Estrogen Status				
Negative	1.00	-	1.00	-
Positive	0.950	0.743, 1.214	0.969	0.755, 1.244
Age				
<40	1.00*	-	1.00	-
40-44	0.868	0.471, 1.597	0.852	0.461, 1.576
45-49	1.820	1.008, 3.284	1.815	1.001, 3.293
50-54	0.648	0.375, 1.121	0.672	0.382, 1.184
55-59	0.624	0.362, 1.078	0.619	0.354, 1.085
60-64	0.660	0.391, 1.115	0.655	0.383, 1.121
65-69	0.792	0.470, 1.334	0.788	0.458, 1.357
FH				
0 FDRs with CRC	1.00	-	1.00	-
1 FDRs with CRC	1.208	0.882, 1.655	1.206	0.879, 1.656
2+ FDRs with CRC	1.754	0.784, 3.925	1.669	0.741, 3.757

a - Adjusted for age, due to important differences in the age distributions of cases and controls that may reflect the sampling design

b - Adjusted for all covariates.

* Unadjusted results for age.

OR, odds ratio; CI, confidence interval; NSAID, non-steroidal anti-inflammatory; BMI, body mass index; FH, family history; FDR, first-degree relative; CRC, colorectal cancer.

ASSESSMENT OF MODEL PERFORMANCE

Likelihood Ratio Tests

With the addition of FH to the model, a statistically significant contribution to the model was observed for the males. For the females, there was no significant contribution observed (Table 13).

TABLE 13 – Likelihood ratio tests for Clinical models

Model	Likelihood Ratio	Δ	df	Δdf	p-value
Male					
Base model	1717.329		17	-	-
With FH	1682.889	34.439	19	2	$3.32E^{-8}$
Female					
Base model	1910.829		15	-	-
With FH	1907.990	2.839	17	2	0.24183

df, degrees of freedom; FH, family history.

Area under the receiver operator characteristics (ROC) curve (AUC)

The AUC values, percentage change in AUC, and significance of change in AUC are presented in Table 14, and ROC graphs comparing models with and without FH in Figures 6a (males) and 6b (females). While there is no standard categorizations of strength of discrimination, the observed AUCs would be considered 'no more than modest' on the basis of Hosmer and Lemeshow's categories .[45]

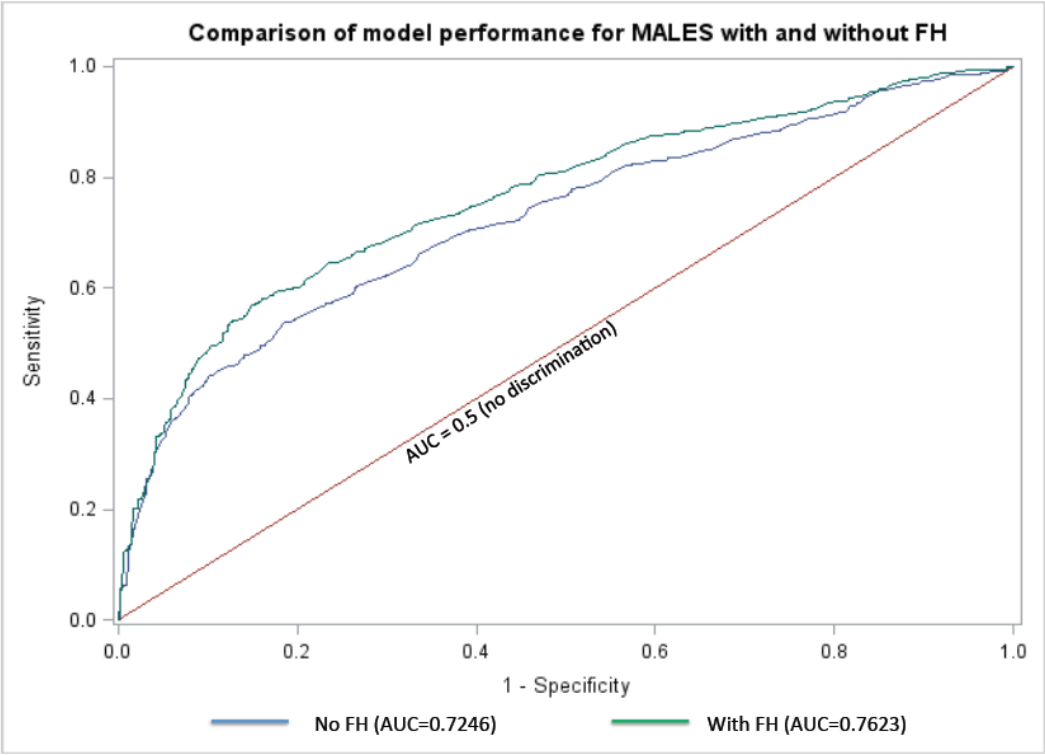


Figure 6a – Male Clinical model comparison

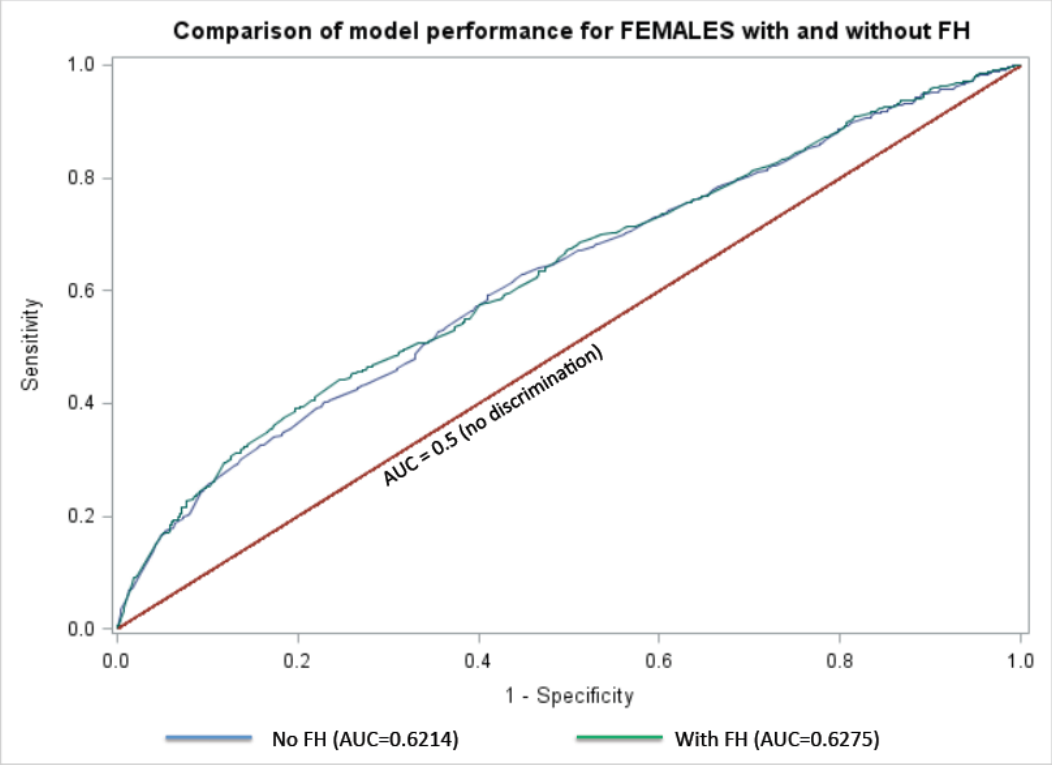


Figure 6b – Female Clinical model comparison

Table 14 – Change in AUCs for base and FH-inclusive Clinical models			
Model	**AUC**	**Δ AUC**	**p-value**
Male			
Base model	0.7246	-	-
With FH	0.7623	+ 0.0037	0.0443
Female			
Base model	0.6214	-	-
With FH	0.6275	+ 0.0061	0.7632

FH, family history; AUC, area under the curve.

The base model of the male sample indicates moderate (acceptable)[45] discriminatory ability. The change in AUCs shows a statistically significant increase in discrimination of + 0.0037 with the addition of FH to the model. The discriminatory ability of the base model for the female sample was marginal, and the addition of FH to the model did not statistically significantly improve discrimination.

2) FAMILY HISTORY MODEL

We compared the age-adjusted ORs and AUCs for FH, stratified by sex, with different definitions of FH (Tables 15a and 15b, Figures 7a and 7b).

Table 15a – Male - OR and AUC when different minimum criteria applied for classification as having 'positive' FH			
Required degree of relationship	**Required number of relatives with CRC**	**OR (95% CI)**	**AUC (95% CI)**
First	≥2	5.451 (1.463, 4.900)	0.7013 (0.6761, 0.7266)
First or second	≥2	6.488 (3.593, 11.714)	0.7415 (0.7174, 0.7656)
First	≥1	2.909 (2.102, 4.025)	0.7443 (0.7204, 0.7682)
First or second	≥1	3.888 (2.948, 5.128)	0.7731 (0.7502, 0.7960)

OR, odds ratio; AUC, area under the curve; FH, family history; CI, confidence interval; CRC, colorectal cancer.

Table 15b – Female - OR and AUC when different minimum criteria applied for classification as having 'positive' FH			
Required degree of relationship	**Required number of relatives with CRC**	**OR (95% CI)**	**AUC (95% CI)**
First	≥2	1.592 (0.760, 3.335)	0.6165 (0.5906, 0.6425)
First or second	≥2	2.619 (1.748, 3.925)	0.6477 (0.6222, 0.6731)
First	≥1	1.300 (0.980, 1.724)	0.6248 (0.5989, 0.6508)
First or second	≥1	1.555 (1.239, 1.952)	0.6442 (0.6185, 0.6699)

OR, odds ratio; AUC, area under the curve; FH, family history; CI, confidence interval; CRC, colorectal cancer.

The highest AUC was used to select the 'best' definition of FH. In men, this was having at least one first or second-degree relative with CRC (AUC=0.773), and, in women, this was having at least two first or second-degree relatives with CRC (AUC=0.6477).

For consistency between male and female models, the definition of 'Having a FH of CRC in at least a second-degree relative' was chosen for the 'best' definition. While it was not the highest performing AUC for women, the difference in model performance (AUC) from the superior model is only 0.0035 less.

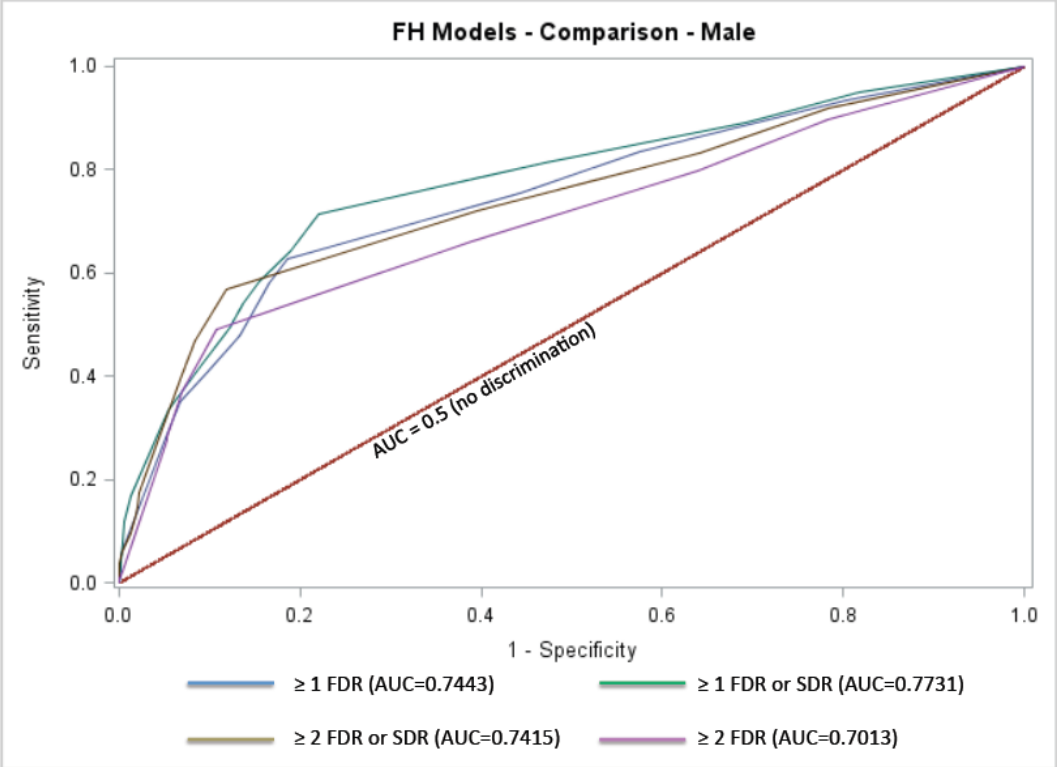


Figure 7a – Male - FH Model performance

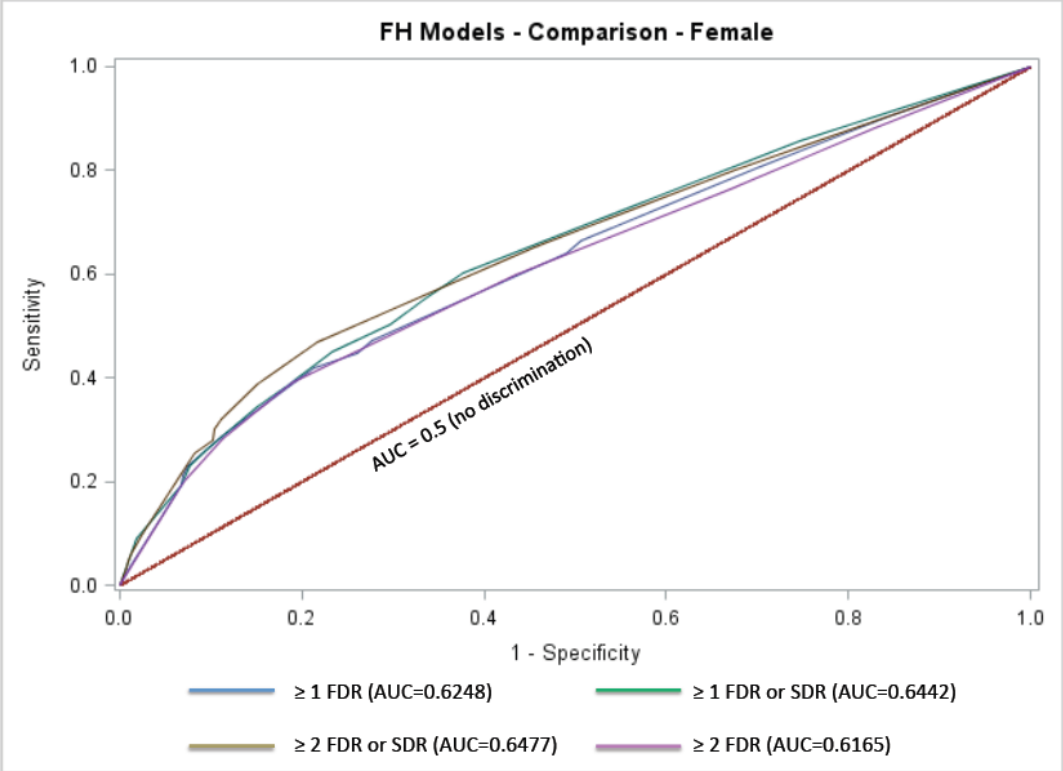


Figure 7b – Female - FH Model performance

3) DATA-DRIVEN (EMPIRICAL) MODEL

In addition to the previously reported findings from the Clinical model (Tables 11a and 11b), results for the variables included only in the Data-driven model are presented in Tables 16a and 16b.

For the males (Table 16a), cases were more likely to have inflammatory bowel disease (4.3% vs. 1.7%), more likely to consume red meat, and less likely to have had post-secondary education (57.9% vs 62.2%) than controls. Cases were more likely to have a positive FH of CRC (36.5%) than controls (13.3%). There were no statistically significant differences between male cases and controls in diabetes or in alcohol consumption.

For the females (Table 16b), cases were more likely to have inflammatory bowel disease (4.3% vs. 2.2%), less likely to be smokers (46.0% vs. 49.4%), and less likely to have post-secondary education (52.5% vs. 60.2%) than controls. Cases were more likely to have a positive FH of CRC (34.7%) than controls (24.6%). There were no statistically significant differences between female cases and controls in diabetes, red meat consumption, or alcohol consumption.

Table 16a - MALE – Comparison of Data-driven model by cases and controls						
Variable	**Division**	**Cases (N=614)**		**Controls (N=664)**		**X^2 Test P-value**
		Un-weighted N (%)	**Weighted %***	**Un-weighted N (%)**	**Weighted %***	
Diabetes	Yes	72 (8.93)	8.60	68 (8.71)	8.71	0.9438
	No	734 (91.07)	91.40	713 (91.29)	91.29	
Inflammatory Bowel Diseases	Yes	36 (4.49)	4.30	13 (1.66)	1.66	0.0048
	No	766 (95.51)	95.70	768 (98.34)	98.34	
Red Meat Consumption	<1 Serving/wk	21 (2.66)	2.30	35 (4.68)	4.68	<0.0001
	1-3 Servings/wk	309 (39.11)	40.66	367 (49.06)	49.06	
	4-6 Servings/wk	244 (30.89)	29.25	196 (26.20)	26.20	
	≥7 Servings/wk	216 (27.34)	27.79	150 (20.05)	20.05	
Alcohol Consumption	0 Drinks/wk	188 (24.48)	23.85	196 (26.06)	26.06	0.0678
	1-2 Drinks/wk	49 (6.38)	5.06	60 (7.98)	7.98	
	3-7 Drinks/wk	167 (21.74)	22.43	162 (21.54)	21.54	
	>7 Drinks/wk	364 (47.40)	48.66	334 (44.41)	44.41	
Education	< High School	184 (22.80)	25.36	191 (24.46)	24.46	0.0445
	High School Grad	136 (16.85)	16.78	104 (13.32)	13.32	
	Some Post-Secondary	270 (33.46)	33.41	249 (31.88)	31.88	
	Post-Secondary Grad	217 (26.89)	24.45	237 (30.35)	30.35	
FH - Any FDRs or SDRs	No	466 (57.11)	63.47	679 (86.72)	86.72	<0.0001
	Yes	350 (42.89)	36.53	104 (13.28)	13.28	

*Weighted proportion to account for sampling design

Counts may not add up to total sample as missing data were excluded.

FH, family history; FDR, first-degree relative; SDR, second-degree relative.

Table 16b - FEMALE – Comparison of Data-driven model by cases and controls						
Variable	**Division**	**Cases (N=706)**		**Controls (N=553)**		**X^2 Test P-value**
		Un-weighted N (%)	**Weighted %***	**Un-weighted N (%)**	**Weighted %***	
Diabetes	Yes	61 (6.12)	6.79	37 (5.17)	5.17	0.1876
	No	936 (93.88)	93.21	679 (94.83)	94.83	
Inflammatory Bowel Diseases	Yes	47 (4.72)	4.28	16 (2.23)	2.23	0.0303
	No	949 (95.28)	95.72	702 (97.77)	97.77	
Red Meat Consumption	<1 Serving/wk	50 (5.34)	5.68	48 (7.05)	7.05	0.3832
	1-3 Servings/wk	461 (49.20)	48.73	344 (50.51)	50.51	
	4-6 Servings/wk	243 (25.93)	25.10	170 (24.96)	24.96	
	≥7 Servings/wk	183 (19.53)	20.49	119 (17.47)	17.47	
Smoking	Never Smoker	513 (53.00)	53.96	347 (50.58)	50.58	0.0113
	>0 and <20 PYs	301 (31.10)	29.85	188 (27.41)	27.41	
	20-29 PYs	58 (5.99)	6.12	44 (6.41)	6.41	
	≥ 30 PYs	96 (9.92)	10.07	107 (15.60)	15.60	
Alcohol Consumption	0 Drinks/wk	529 (57.13)	57.82	381 (56.44)	56.44	0.4879
	1-2 Drinks/wk	96 (10.37)	10.12	59 (8.74)	8.74	
	3-7 Drinks/wk	156 (16.85)	17.02	134 (19.85)	19.85	
	>7 Drinks/wk	145 (15.66)	15.03	101 (14.96)	14.96	
Education	< High School	218 (21.87)	23.44	129 (18.07)	18.07	0.0163
	High School Grad	224 (22.47)	24.09	155 (21.71)	21.71	
	Some Post-Secondary	345 (34.60)	32.66	269 (37.68)	37.68	
	Post-Secondary Grad	210 (21.06)	19.81	161 (22.55)	22.55	
FH - Any FDRs or SDRs	No	610 (60.46)	65.26	543 (75.42)	75.42	<0.0001
	Yes	399 (39.54)	34.74	177 (24.58)	24.58	

*Weighted proportion to account for sampling design

Counts may not add up to total sample as missing data were excluded.

PYs, pack years; FH, family history; FDR, first-degree relative; SDR, second-degree relative.

REGRESSION MODELING

Model Selection

Two pairs of models (Tables 17a, 17b) were run in a data-driven manner. For both males and females, base models were created without FH. All candidate variables were entered into the regression analysis, and final models include only significant (at $p<0.05$) predictor variables determined through backward elimination method for stepwise regression.

Candidate Models

Table 17a – MALE – Data-driven model eliminated and significant variables

Variable	P-value
Eliminated Variables	
Diabetes	0.7406
Smoking	0.6671
Physical Activity	0.6007
Vegetable Consumption	0.5728
Education	0.3773
Alcohol Consumption	0.1512
Inflammatory Bowel Disease	0.0996
Significant Predictors	
Age	<0.0001
BMI	0.0001
NSAID Use	0.0450
Red Meat Consumption	0.039

NSAID, Non-steroidal anti-inflammatory; BMI, body mass index.

Table 17b – FEMALE – Data-driven model eliminated and significant variables

Variable	P-value
Eliminated Variables	
Vegetable Consumption	0.9701
BMI	0.9471
Diabetes	0.8111
Red Meat Consumption	0.6830
Estrogen Status	0.6061
Alcohol Consumption	0.3719
Physical Activity	0.2784
Smoking	0.0544
Significant Predictors	
Age	0.0055
Inflammatory Bowel Diseases	0.0485
NSAID Use	0.0008
Education	0.0335

The results of the multivariate models, including FH, are presented in Tables 18a and 18b. Age-adjusted ORs are presented, to account for age differences that were apparent from the sampling design, and multivariate-adjusted ORs are presented for each variable included in the model.

Males

For the male population (Table 18a), the odds of CRC are statistically significantly increased for those in the ‘overweight’ and ‘obese’ categories, and for those who consume greater quantities of red

meat. Odds were significantly reduced for those who regularly used NSAID medication. As compared to those with no FH of CRC, the odds are significantly increased for those with a positive FH (OR=4.03).

Table 18a – Male – Age and Multivariate adjusted ORs for Data-driven model

Variable	Division	Age-Adjusted OR[a]		Multivariate-Adjusted OR[b]	
		OR	95% CI	OR	95% CI
BMI	Under/Normal	1.00	-	1.00	-
	Overweight	1.713	1.308, 2.243	1.746	1.316, 2.317
	Obese	1.935	1.371, 2.732	1.850	1.289, 2.655
Red Meat Consumption	<1 Serving/wk	1.00	-	1.00	-
	1-3 Servings/wk	1.865	0.984, 2.533	1.754	0.909, 3.385
	4-6 Servings/wk	2.593	1.349, 4.983	2.239	1.142, 4.389
	≥7 Servings/wk	2.785	1.438, 5.392	2.567	1.298, 5.074
NSAID use	No	1.00	-	1.00	-
	Yes	0.598	0.354, 1.010	0.470	0.270, 0.816
Age	<40	1.00*	-	1.00	-
	40-44	0.556	0.232, 1.330	0.592	0.239, 1.468
	45-49	1.164	0.499, 2.718	1.365	0.569, 3.276
	50-54	0.238	0.110, 0.516	0.255	0.114, 0.569
	55-59	0.189	0.088, 0.404	0.184	0.084, 0.406
	60-64	0.272	0.129, 0.575	0.298	0.137, 0.646
	65-69	0.261	0.124, 0.550	0.284	0.131, 0.616
FH - Any FDRs or SDRs	No	1.00	-	1.00	-
	Yes	3.895	2.878, 5.271	4.034	2.957, 5.502

a - Adjusted for age, due to important differences in the age distributions of cases and controls that may reflect the sampling design

b - Adjusted for all covariates.

* Unadjusted results for age.

OR, odds ratio; CI, confidence interval; NSAID, non-steroidal anti-inflammatory; BMI, body mass index; FH, family history; FDR, first-degree relative; CRC, colorectal cancer.

Females

For the female population (Table 18b), the odds are significantly increased for those with inflammatory bowel diseases, and significantly decreased for those who regularly use NSAID medication and those who have a post-secondary education. As compared to those with no FH of CRC, there is a significant increase in odds for those with any affected first- or second-degree family members (OR=1.77).

Table 18b – Female – Age and Multivariate adjusted ORs for Data-driven model					
Variable	**Division**	**Age-Adjusted OR[a]**		**Multivariate-Adjusted OR[b]**	
		OR	**95% CI**	**OR**	**95% CI**
NSAID use	No	1.00	-	1.00	-
	Yes	0.447	0.279, 0.718	0.478	0.307, 0.745
Inflammatory Bowel Diseases	No	1.00	-	1.00	-
	Yes	2.003	0.996, 4.030	2.089	1.023, 4.264
Education	< High School	1.00	-	1.00	-
	High School Grad	0.794	0.545, 1.156	0.767	0.523, 1.123
	Some Post-Secondary	0.609	0.434, 0.853	0.590	0.418, 0.832
	Post-Secondary Grad	0.617	0.420, 0.906	0.609	0.412, 0.901
Age	<40	1.00*	-	1.00	-
	40-44	1.027	0.532, 1.983	1.114	0.572, 2.172
	45-49	1.705	0.919, 3.163	1.600	0.852, 3.005
	50-54	0.681	0.383, 1.210	0.662	0.368, 1.190
	55-59	0.731	0.411, 1.299	0.731	0.406, 1.319
	60-64	0.792	0.457, 1.374	0.794	0.450, 1.400
	65-69	0.938	0.542, 1.621	0.892	0.505, 1.575
FH – Any FDRs or SDRs	No	1.00	-	1.00	-
	Yes	1.706	1.313, 2.217	1.767	1.354, 2.304

a - Adjusted for age, due to important differences in the age distributions of cases and controls that may reflect the sampling design

b - Adjusted for all covariates.

* Unadjusted results for age.

OR, odds ratio; CI, confidence interval; NSAID, non-steroidal anti-inflammatory; BMI, body mass index; FH, family history; FDR, first-degree relative; CRC, colorectal cancer.

ASSESSMENT OF DATA-DRIVEN MODEL PERFORMANCE

Likelihood Ratio Tests

With the addition of FH to the model, a statistically significant contribution to the model was observed for both models (Table 19).

Table 19 – Likelihood ratio tests for Data-driven models					
Model	**Likelihood Ratio**	**Δ**	**df**	**Δdf**	**p-value**
Male					
Base model	1595.890		13	-	-
With FH	1506.560	89.33	14	1	<0.0001
Female					
Base model	1598.070		12	-	-
With FH	1579.833	18.237	13	1	<0.0001

df, degrees of freedom; FH, family history.

AUC comparison of Data-driven models

The discriminatory ability of the models was assessed using the AUC. AUC values, percentage change, significance of change are presented in Table 20, and ROC graphs comparing models with and without FH are presented in figures 8a (males) and 8b (females).

Table 20 – Change in AUCs for base and FH-inclusive Data-driven models

Model	AUC	Δ AUC	p-value
Male			
Base model	0.7250	-	-
With FH	0.7931	+ 0.0681	0.0003
Female			
Base model	0.6165	-	-
With FH	0.6475	+ 0.0310	0.1605

FH, family history; AUC, area under the curve.

The base model of the male sample indicates moderate (acceptable)[45] discriminatory ability. The change in AUCs shows a statistically significant increase in discrimination of + 0.0681 with the addition of FH to the model. The discriminatory ability of the base model for the female sample was marginal, and an increase in AUC of +0.0310 is not statistically significant.

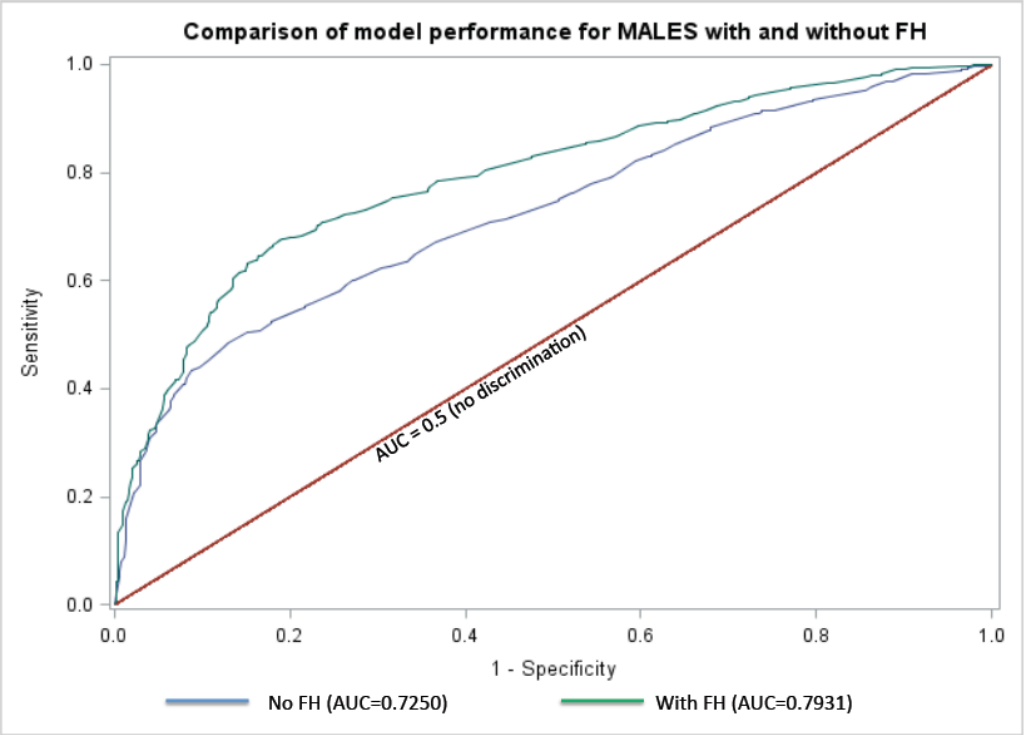


Figure 8a – Male - Data-driven model comparison

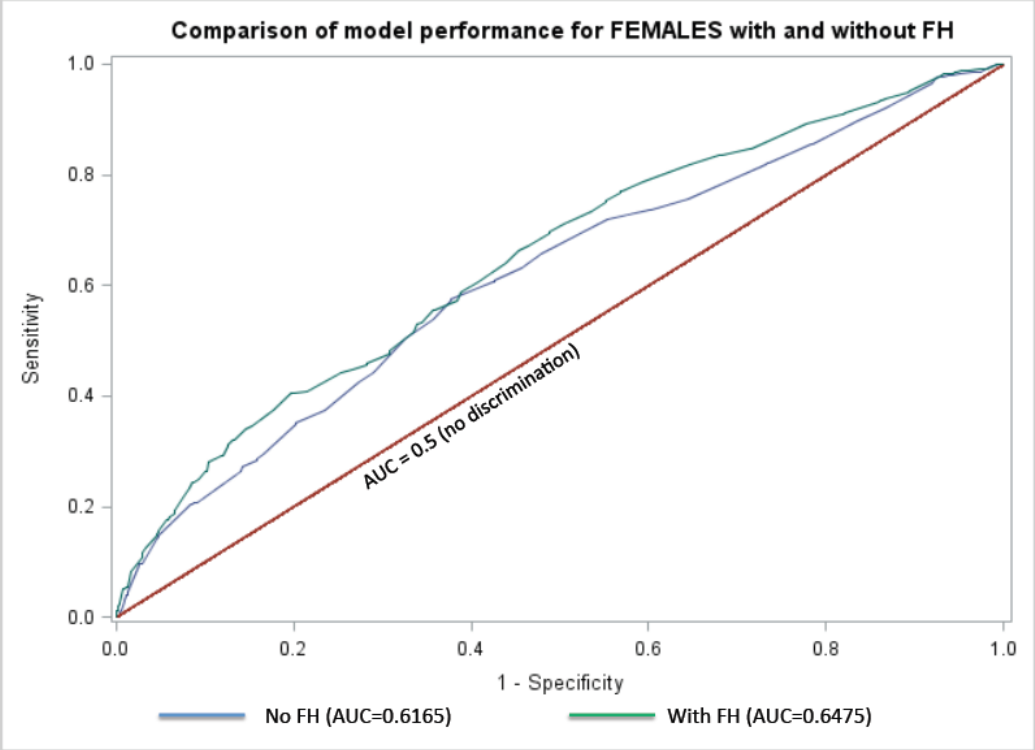


Figure 8b – Female - Data-driven model comparison

4) STRATIFIED AND SENSITIVITY ANALYSES

STRATIFIED SUB-ANALYSIS

Due to lack of sufficient data in the <40 age group, the intended age-stratified analysis was not undertaken for the Clinical model due to a small sample size in the <40 age group. The analysis was restricted to the ≥40 group. Frequency tables for the <40 age group, multivariate-adjusted ORs for the ≥40 group, and comparison of AUCs of the ≥40 and the all ages groups are provided in Appendix F. Results for the ≥40s group were similar to those of the full sample.

SENSITIVITY ANALYSES

Polyp and Screening History

The 'Polyp and Screening History' variable was removed from the model because we were concerned that its inclusion would conflate its status as a risk factor with its status as a marker of cancer detection. An analysis was run with the variable re-introduced to test the change in model performance with its inclusion. Results of the re-inclusion of the Freedman's polyp/screening variable are shown in Tables 21a, 21b and 22.

The inclusion of the Polyp and Screening History variable into the model had a very large effect on the model. Its inclusion controlled the model, with FH becoming statistically non-significant (Tables 21a and 21b). The change in AUC with the addition of FH was +0.0034 in males and +0.0007 in females (Table 22).

Table 21a – MALE – Clinical model multivariate-adjusted ORs – Models with and without Polyp and Screening History		
	With Polyp/Screening OR (95% CI)	**Without Polyp/Screening OR (95% CI)**
BMI		
Under/Normal	1.00	1.00
Overweight	1.85 (1.31, 2.63)	1.89 (1.45, 2.48)
Obese	2.42 (1.52, 3.86)	2.27 (1.61, 3.19)
Physical Activity		
0 hrs/week	1.00	1.00
>0 and ≤2 hrs/week	1.21 (0.64, 2.31)	1.36 (0.83, 2.23)
>2 and ≤4 hrs/week	0.85 (0.44, 1.64)	0.98 (0.58, 1.64)
>4 hrs/week	1.34 (0.78, 2.29)	1.14 (0.75, 1.71)
NSAID Use		
No	1.00	1.00
Yes	0.46 (0.23, 0.90)	0.55 (0.33, 0.93)
Vegetables		
< 5 Servings/day	1.00	1.00
≥5 Servings/day	1.72 (0.46, 6.44)	1.55 (0.52, 4.62)
Smoking		
Never Smoker	1.00	1.00
>0 and <20 PYs	1.08 (0.74, 1.59)	1.22 (0.91, 1.62)
20-29 PYs	0.93 (0.56, 1.56)	1.38 (0.93, 2.05)
≥ 30 PYs	0.95 (0.62, 1.46)	1.23 (0.89, 1.68)
Age		
<40	1.00	1.00
40-44	0.44 (0.13, 1.46)	0.58 (0.24, 1.39)
45-49	1.11 (0.36, 3.41)	1.02 (0.44, 2.37)
50-54	0.14 (0.05, 0.40)	0.21 (0.09, 0.46)
55-59	0.09 (0.03, 0.26)	0.15 (0.07, 0.34)
60-64	0.11 (0.04, 0.32)	0.23 (0.11, 0.49)
65-69	0.09 (0.03, 0.24)	0.22 (0.10. 0.47)
FH		
0 FDRs with CRC	1.00	1.00
1 FDR with CRC	1.31 (0.83, 2.07)	2.42 (1.66, 3.51)
≥2 FDRs with CRC	2.89 (0.83, 10.04)	5.51 (1.81, 16.72)
Polyp/Screening		
No Sigmoidoscopy/Colonoscopy	1.00	-
Sigmoidoscopy/Colonoscopy and no HX Polyps	39.71 (25.05, 62.94)	
Sigmoidoscopy/Colonoscopy and HX of Polyps	108.54 (63.98, 184.15)	
Sigmoidoscopy/Colonoscopy and HX Polyps Unknown	20.87 (10.32, 42.21)	

OR, odds ratio; CI, confidence interval; BMI, body mass index; PYs, pack years; FH, family history; FDR, first-degree relative; SDR, second-degree relative; HX, history.

Table 21b– FEMALE – Clinical model multivariate-adjusted ORs – Models with and without Polyp and Screening History		
	With Polyp/Screening OR (95% CI)	**Without Polyp/Screening OR (95% CI)**
BMI		
Under/Normal	1.00	1.00
Overweight	1.02 (0.72, 1.46)	1.12 (0.87, 1.44)
Obese	0.78 (0.51, 1.18)	1.06 (0.78, 1.43)
Physical Activity		
0 hours/week	1.00	1.00
>0 and ≤2 hours/week	0.73 (0.38, 1.42)	1.38 (0.88, 2.18)
>2 and ≤4 hours/week	0.86 (0.44, 1.64)	1.36 (0.88, 2.09)
>4 hours/week	0.86 (0.50, 1.48)	1.46 (1.01, 2.10)
NSAID Use		
Yes	0.51 (0.28, 0.91)	0.48 (0.31, 0.75)
No	1.00	1.00
Vegetables		
< 5 Servings/day	1.00	1.00
≥5 Servings/day	0.80 (0.43, 1.49)	0.81 (0.51, 1.29)
Estrogen Status		
Negative	1.00	1.00
Positive	0.78 (0.55, 1.11)	0.97 (0.76, 1.24)
Age		
<40	1.00	1.00
40-44	0.90 (0.36, 2.27)	0.58 (0.24, 1.39)
45-49	1.59 (0.66, 3.84)	1.02 (0.44, 2.37)
50-54	0.47 (0.20, 1.10)	0.21 (0.09, 0.46)
55-59	0.31 (0.13, 0.70)	0.15 (0.07, 0.34)
60-64	0.30 (0.14, 0.68)	0.27 (0.11, 0.49)
65-69	0.30 (0.14, 0.68)	0.22 (0.10, 0.47)
FH - FDRs with CRC		
0 FDRs with CRC	1.00	1.00
1 FDR with CRC	0.69 (0.45, 1.06)	1.21 (0.88, 1.66)
≥2 FDRs with CRC	1.05 (0.37, 2.99)	1.67 (0.74, 3.76)
Polyp/Screening		
No Sigmoidoscopy/Colonoscopy	1.00	-
Sigmoidoscopy/Colonoscopy and no HX Polyps	44.86 (29.25, 68.82)	
Sigmoidoscopy/Colonoscopy and HX of Polyps	187.77 (110.57,	
Sigmoidoscopy/Colonoscopy and HX Polyps	318.85)	
Unknown	22.24 (11.36, 43.56)	

OR, odds ratio; CI, confidence interval; BMI, body mass index; FH, family history; FDR, first-degree relative; SDR, second-degree relative; HX, history.

Table 22 –Comparison of AUCs for models with and without Polyp and Screening History variable		
Model	**AUCs Without Polyp/Screening**	**AUCs With Polyp/Screening**
Male		
Base **FH - Included**	0.7246 0.7623	0.9004 0.9038
Δ AUC	+ 0.0377	+ 0.0034
P-value	0.0443	0.7619
Female		
Base **FH - Included**	0.6214 0.6275	0.8973 0.8980
Δ AUC	+ 0.0061	+0.0007
P-value	0.7632	0.9508

FH, family history; AUC, area under the curve.

Estrogen Status

In order to examine our coding of the estrogen status variable, we re-ran the multivariable analysis excluding female participants +/- 3 years of the average menopause age in Canada (51 years), since participants in this age group were the most likely to have been misclassified with respect to estrogen status. Differences in ORs for estrogen status as well as changes in AUC are presented in Table 23.

Table 23– Changes in model performance based on Estrogen Status definition		
	Clinical model (original) definition	Excluding participants aged 48-54 years
OR - Estrogen Status	0.969 (0.755, 1.244)	1.076 (0.814, 1.422)
AUC Base FH-Included	 0.6214 0.6275	 0.6118 0.6206
Δ AUC	+ 0.0061	+ 0.0088
P-value	0.1605	0.6954

OR, odds ratio; FH, family history; AUC, area under the curve.

With the exclusion of participants 48-54 years of age, very little change in AUC with FH was observed compared with the primary analysis.

Identification of Participants with Lynch syndrome

Participants who were coded as definitely fulfilling the Amsterdam criteria for Lynch syndrome were excluded prior to analysis. Participants coded as probably fulfilling the Amsterdam criteria (n=29) were further excluded to determine whether FH appeared to have an effect on the AUC for sporadic forms of CRC. Changes for FH ORs and changes in the AUC are presented in Tables 24a and 24b.

For both males and females, excluding such participants had little effect on the absolute change in the AUC when FH was added to the base model, however the change in AUC for males no longer remained significant (p=0.1292).

Table 24a – MALE – Changes in model performance based on Amsterdam definition		
	Clinical model (original) definition	Excluding participants fulfilling definite or probable Amsterdam criteria
OR – FH		
1 FDR with CRC	2.42 (1.66, 3.51)	2.27 (1.56, 3.31)
2+ FDR with CRC	5.51 (1.81, 16.72)	3.64 (1.15, 11.51)
AUC		
Base	0.7246	0.7248
FH-Included	0.7623	0.7539
Δ AUC	+ 0.0377	+ 0.0291
P-value	0.0443	0.1292

OR, odds ratio; FH, family history; CRC, colorectal cancer; AUC, area under the curve.

Table 24b – FEMALE – Changes in model performance based on Amsterdam definition		
	Clinical model (original) definition	Excluding participants fulfilling definite or probable Amsterdam criteria
OR – FH		
1 FDR with CRC	1.21 (0.88, 1.66)	1.01 (0.73, 1.39)
2+ FDR with CRC	1.67 (0.74, 3.76)	1.31 (0.56, 3.06)
AUC		
Base	0.6214	0.6107
FH-Included	0.6275	0.6116
Δ AUC	+ 0.0061	+ 0.0009
P-value	0.7632	0.9628

OR, odds ratio; FH, family history; CRC, colorectal cancer; AUC, area under the curve.

CHAPTER IV - DISCUSSION

The 2009 NIH State-of-the-science panel established the need for evidence to support the use of FH information for common complex diseases such as CRC.[12] The purpose of this project was to improve the underlying evidence base regarding the clinical validity of FH in CRC, specifically its value in risk classification. This project investigated whether the addition of a FH variable to regression models designed to classify CRC risk improved their performance. Additionally, differing definitions of FH of CRC were investigated to identify the definition with the best performance. Evidence regarding clinical validity of FH will aid in informing clinical utility analyses.

We developed two models, one based on the existing NCI ("Freedman") tool [30] (the 'clinical model') and one which used all the variables available to us in the ARCTIC dataset (the 'data driven model'). Clinical model analyses indicated that BMI, NSAID use, and FH were independent predictors of CRC for males, and that age and NSAID use were independent predictors of CRC for females.

Overall, the core clinical models (without FH) were able to correctly classify about two-thirds to three-quarters of cases and controls to their correct status (true positives and true negatives). Even though the association between FH and case/control status was large, as judged by the odds ratios, the change in accuracy of risk classification was very slight (with absolute gains of less than 5% for males and 1% for females), even if statistically significant for males. Our data-driven model analysis indicated that age, BMI, NSAID use, consumption of red meat, and FH were independent predictors of CRC for males, and that age, NSAID use, inflammatory bowel disease, education, and FH were significant predictors for females.

Similarly to the clinical models, the data-driven models were able to correctly classify two-thirds to three quarters of the participants to their correct disease status. The addition of FH to the base

models created larger changes in risk classification (6.8% males, 3.1% females) compared with the clinical models, however, this change remained statistically significant for males only. The ORs for FH were significant for both men and women, with values of 4.0 and 1.8 respectively.

The lack of statistically significant change in model discrimination for females, with the inclusion of FH, is surprising. In the clinical model, there was no difference between cases and controls in terms of having a positive family history; in that case, FH was not a significant predictor of CRC risk. In the data-driven model, there was a significant difference between cases and controls in having a positive FH of CRC, however, the magnitude of the association (ORs) was not great (as compared to the males). It is possible that the magnitude of the effect of FH was not sufficient to significantly improve model discrimination.

Thus, while the addition of FH to the models adds little to disease prediction for women, there is significant improvement in risk prediction for males in both the clinical and data-driven models. While the improvements to model performance were slight, FH remained an independent risk predictor of CRC. The minimal increase in model discrimination may indicate that FH may not be a useful risk factor for improving CRC risk models in this dataset.

From the FH modeling analyses, the "best" definition of FH was having at least one first- or second-degree relative with CRC. This is the broadest of the tested definitions, and the one for which most participants are likely to have a positive history.

The model discrimination and improvements in risk classification determined through the clinical and data-driven modeling are consistent with findings from other assessments of FH in complex disease. Our model discrimination (AUCs) were comparable, if not slightly better (for males) to what has been achieved in a recently published CRC risk model replication study (which includes FH), which reported an AUC of 61 percent for men and women.[31] It is also comparable to risk models in other

cancers (breast, lung, ovarian, melanoma)[31] and cardiovascular disease,[46] which generated AUCs ranging from 60 to 69 percent. The improvements in model discrimination, with the inclusion of FH, that were found are comparable in magnitude to those found in an assessment of the added values of FH in cardiovascular disease and coronary heart disease.[46–48]

Other than FH, there was another exposure that had a statistically significant effect in all models. Regular NSAID use had a statistically significant protective effect. This finding is consistent with published evidence regarding the effect of NSAID medication in CRC risk.[27,30,43,49,50] While the mechanism of this protective effect is not entirely understood, the main arguments currently suggest that NSAIDs reduce cell growth or inhibit tumor growth through altering insulin resistance.[50,51]

Strengths and Limitations

This study had both strengths and limitations. ARCTIC is a large dataset derived from the OFCCR [32,33] with a broad range of risk factor data and the necessary FH information. The OFCCR has been frequently used for investigation of various risk factors as well as genome-based studies.[43,52–55] There were three major strengths in using this dataset. The first was the availability of a large dataset for the planned analyses.

Secondly, the ARCTIC dataset questionnaires had detailed assessment of FH. This detailed information permitted the potential creation of multiple definitions of FH, which we were able to create and compare through regression analyses to identify the top performing FH definition for this dataset.

The third major advantage was the careful ascertainment of cases in the original study through pathology records, meaning that misclassification of case status was unlikely.

However, working with a pre-existing dataset presented many challenges. The first was related to the unequal probability of sampling cases to participate in the OFCCR dataset. The oversampling of higher risk individuals meant that cases with a FH of CRC were more likely to be selected to participate

in the OFCCR than cases without a FH. This had the potential to create serious bias for our analysis of the association between FH and CRC (see Methods, Table 1). A mitigating factor was that all cases were sampled with a *known* probability so that the original study investigators were able to generate sampling weights to account for the unequal probability of selection of cases. Given the objectives of this thesis, this issue was discussed extensively within the Thesis Advisory Committee, and with biostatisticians familiar with the OFCCR data. The conclusion was that applying the appropriate sampling weights would limit any effect of the unequal probability of selection of cases related to FH.

A second challenge, common to secondary analysis studies, was that our analysis was constrained by the variables and measurements used by the OFCCR. While a broad range of variables were available, some characteristics that we would have preferred to include in the clinical model (for comparability with Freedman's work)[30] were not available, not complete, or defined in a way quite different from the definitions used by Freedman and colleagues. This required the re-construction of variables based on the information available in the dataset, particularly relating to physical activity and alcohol consumption. For example, in the ARCTIC study, participants were asked to recall their physical activity levels and alcohol consumption from previous decades in lengthy series' of questions and item non-response resulted in summary variables with high levels of missing data.

All methods of dealing with missing data have limitations and may introduce bias. The potential for bias increases as the proportion of participants with missing data increases. In our dataset, we were most concerned about missing data for physical activity and alcohol consumption, which, upon initial creation, were missing in 16-22% of participants. The participant exclusion (also known as casewise deletion) approach would likely have introduced selection bias had we relied exclusively on this method for alcohol consumption and physical activity because those who had complete data may have differed from those with missing data; this method would also have substantially reduced our sample size for analysis.[56,57] The variable deletion approach was not an option because it would have removed an

important predictor variable. We therefore chose to use a form of imputation based on making logical assumptions about the consistency of lifestyle habits over time, carrying forward values from an earlier decade when recent values were missing. Our imputation method assumed that the respondent's report of the previous decade's physical activity or alcohol consumption was representative of current use. If this assumption was incorrect for some participants, it would have led to some degree of exposure misclassification. We also appreciate that this method still carried the risk of differential missingness between cases and controls, given that the reasons for missingness were not known.[56,57]

The risk factor data (including FH) were collected through questionnaires. Self-report data have reported accuracy rates between 65-85%,[58] dependent on both the population and the constructs measured. Reviews of the accuracy of FH reporting indicate generally high specificity (correctly reporting absence of disease in relatives) but more modest sensitivity (correctly reporting affected relatives).[59] Using self-report data carries the risk of introducing bias. For example, self-report data regarding BMI (measures of weight and height) are shown to be biased, as compared to direct measures of weight and height – prevalence of obesity was lower when calculations were based on self-report data.[60] All self-report data carries the risk of inaccuracy of recall, particularly for variables that require recall of activities in previous decades, such as alcohol consumption and physical activity.

More likely to have an effect on the validity of our results is the potential for recall bias (differential recall) of exposures between cases and controls. It has been noted that cases are suspected of providing a more complete report of their true exposure to a risk factor, such as FH, [61] because of the personal experience of a serious illness. If cases were more likely to recall a positive FH when present, relative to controls, this would bias the estimate of the odds ratio for FH as a risk factor for CRC upward, i.e., it would inflate the estimated positive association between FH and CRC. [61] However, in our study, the odds ratios for FH as a risk factor for CRC were comparable to those found in a systematic review of

effect of FH in CRC, which diminishes our concern about the potential for recall bias to account for the observed association. [6]

While there was frequency matching by age (5-year age categories) and sex for cases and controls, there was still mismatch in age distribution between the two groups. All models included age to control for this. However the estimates of the association of age with CRC cannot easily be interpreted as an exposure variable as age was part of participant selection criteria. This mismatch in age-distribution could be due to differential response rates of participants -the frequency matching could have occurred at the sampling stage, and then older controls were those who were more likely to participate. The mismatch may also have resulted from the sample reduction process that was used to limit the impact of unequal probabilities of case selection in the OFCCR. Due to the mismatch, model data regarding the effect of age was not interpreted.

The selection of reference groups for regression models is important, as the choice of reference group can make estimates of association more difficult to interpret. For example, in the analyses conducted, the reference group of 0 drinks/week was used for the alcohol consumption variable. This group contains those who have never consumed alcohol and those who have stopped consuming it (possibly for health reasons). There may be differences in risk between these two groups represented in the 0 drinks/week category. In the case of these analyses, the reference categories for included variables were chosen corresponding to those used in the Freedman model, or those used in other published analyses of CRC risk. In further analyses, the low consumption group (1-2 drinks/week), could be considered for use as the reference group.

We were sufficiently concerned about the structure or definition of some variables that we conducted sensitivity analyses to examine the robustness of our results. Firstly, we were aware of likely differences in screening approaches and uptake between the population in which the Freedman model

was developed and our own study population, differences that might mean that the polyp and screening history variable would reflect more than just presence of pre-cancerous lesions. Sigmoidoscopy and colonoscopy are used more frequently as primary screening modalities (for general, asymptomatic, populations) in US healthcare than in Canada, where they are more likely to be used for next stage screening or diagnostic investigation because of symptoms or positive FOBT results.[22,62] If so, this variable would be more likely to act as a marker of higher prior probability of disease and introduce selection (detection) bias for an analysis within a Canadian sample. The sensitivity analysis supported this concern, in that the variable became dominant in the model, with FH becoming non-significant as a predictor. We concluded that, in this dataset, polyp and screening history indeed indicated detection bias and its exclusion from the model was warranted. While Polyp and Screening history were removed from the model, it is important to note that a history of polyps and history of colorectal screening are important predictors of CRC risk and important components of risk assessment and management.[22,43] These factors are important for consideration for inclusion when developing risk assessment models.

The second sensitivity analysis was prompted because our dataset contained no definitive marker of estrogen status, which was a contributing predictive factor in Freedman's clinical model. Rather, we had information on age and on use of hormone replacement medication, which we used to create a composite variable that we hoped would be a reasonable substitution. Without individual data on menopausal status, we had to use the average age of menopause in Canada as a proxy cut-off point, so we were concerned about potential misclassification. The sensitivity analysis indicated little change to the OR estimates (remained non-significant) when the proportion of the sample at greatest risk of being misclassified was removed, providing some confidence that our composite variable was a reasonable measure of estrogen status.

Finally, in our consideration of CRC risk assessment, we wished to account for the fact that there are known genetic forms of the disease. The intention of the study was to gauge the value of FH as a

predictor relevant to the majority, so-called 'sporadic', forms of CRC, rather than the two specific syndromes, FAP and Lynch syndrome. These conditions would meet our simple 'positive FH' definitions used in our models, but more extensive FH evidence is required for their diagnosis. Variables in the ARCTIC dataset identified cases of FAP. There were also variables that identified both confirmed and probable Lynch syndrome cases. We removed confirmed cases prior to all analysis, and wished to check the effect of a stricter exclusion policy (i.e., that our sample was as close as possible to only 'sporadic' cases of CRC) by further excluding probable cases of Lynch syndrome. These exclusions further reduced the change in AUC, in the clinical models, by 0.86%, however the change in AUC with the addition of FH was no longer statistically significant (p=0.129). The small magnitude of the difference in the AUC change with these exclusions suggests that the presence of some cases of Lynch syndrome in the case sample is unlikely to account for the positive association we observed between FH and CRC in our main analyses. While the AUC change did become non-significant, this could partially be due to the decrease in sample size.

All the analyses in this project made use of fairly simple definitions of a positive FH. Definitions of FH used in chronic disease and cancer-specific studies frequently include number of affected family members as well as their age at diagnosis. [21,63–68] It is possible that the predictive accuracy of FH might be improved if stricter criteria were applied, such as a diagnosis in a relative before a defined age and/or requiring two or more affected relatives. These would have the effect of increasing specificity. [22] Evaluating different FH criteria requires access to datasets that contain the full range of FH data for each individual as well as all other risk factor data.

The models used in these analyses have not been validated. Replication of these results is an important next step to develop a more robust evidence base; however this step is made difficult by the need to secure sufficiently large datasets that contain sufficient risk factor, FH, and disease outcome variables. Replication of the models in datasets derived from cohort studies or longitudinal analyses

would not only allow validation of these models, but would also allow for risk re-classification analysis. Risk re-classification analysis was not possible with case-control data, due to the baseline probability of disease being established by the study design (approximately 50%). Following a regression analysis with risk re-classification analyses would provide additional evidence on the clinical utility of FH in CRC risk assessment.

Conclusions and Implications

In this study, we demonstrated that adding a FH variable to an existing CRC risk assessment model (developed by Freedman and colleagues) appears to improve risk classification for males but not females. In a second model developed using all available variables in our dataset, the risk classification improved, with the addition of FH, to a greater degree, but similar to the clinical model, this increase was statistically significant only for males. However, in all models where improvement was observed, the absolute change in AUC was small.

The clinical relevance of these increases in predictive accuracy depends on the changes in preventive and clinical management that follow. Risk re-classification analyses would provide further insight, including whether any improved performance of a model was related to reductions in false negatives, false positives, or both. A net reduction in false negatives implies better detection of those at genuinely increased risk, and enhanced surveillance activities. In contrast, a net reduction in false positive implies avoidance of unnecessary surveillance interventions in those who are not at increased risk. Risk re-classification analysis requires that risk cut-points and a minimal clinically significant difference be firmly established prior to completing the analysis so as to be able to determine whether an important change has occurred.

Taking FH is a core element of health professional practice, but is an activity with opportunity cost at least in terms of professional time. Screening of patients through colonoscopy also carries costs

to the healthcare system, and risks to the patient being screened. Using the empirical data from risk re-classification analysis in decision analytic modeling, for example, could help quantify the clinical utility of incorporating FH into screening and preventive efforts, promoting evidence-based use of scarce health resources.

REFERENCES

1. Davis J. Millions to be poured into personalized health care - Plan tailors treatment to individual's needs. *Ottawa Citizen*. http://www.ottawacitizen.com/health/Millions+poured+into+personalized+health+care/6075695/story.html. Published January 31, 2012.

2. Conti R, Veenstra DL, Armstrong K, Lesko LJ, Grosse SD. Personalized medicine and genomics: challenges and opportunities in assessing effectiveness, cost-effectiveness, and future research priorities. *Med. Decis. Making*. 2010;30(3):328–40.

3. Ginsburg GS, McCarthy JJ. Personalized medicine: revolutionizing drug discovery and patient care. *Trends Biotechnol.* 2001;19(12):491–6.

4. Wilson BJ, Carroll JC, Allanson J, et al. Family history tools in primary care: does one size fit all? *Public Health Genomics*. 2012;15(3-4):181–8.

5. Wilson BJ, Qureshi N, Santaguida P, et al. Systematic Review : Family History in Risk Assessment for Common. *Ann. Intern. Med.* 2009;151(12):878–885.

6. Butterworth A. Family history as a risk factor for common , complex disease. *Public Heal. Genet.* 2007.

7. Yoon PW, Scheuner MT, Peterson-Oehlke KL, Gwinn M, Faucett A, Khoury MJ. Can family history be used as a tool for public health and preventive medicine? *Genet. Med.* 2002;4(4):304–10.

8. Yoon PW, Scheuner MT, Khoury MJ. Research priorities for evaluating family history in the prevention of common chronic diseases. *Am. J. Prev. Med.* 2003;24(2):128–135.

9. Qureshi N, Wilson B, Santaguida P. Family History and Improving Health. *AHRQ - Evid. Report/Technology Assess.* 2009;(186).

10. Wilson B, Qureshi N, Little J, et al. Clinical utility of cancer family history collection in primary care. *Evid. Rep. Technol. Assess. (Full. Rep).* 2009;(179):1–94.

11. Hariri S, Yoon PW, Qureshi N, Valdez R, Scheuner MT, Khoury MJ. Family history of type 2 diabetes: A population-based screening tool for prevention? *Genet. Med.* 2006;8(2):102–108.

12. Berg AO, Baird MA, Botkin JR, Driscoll DA, Fishman PA. National Institutes of Health State-of-the-Science Conference Statement : Family History and Improving Health. *Ann. Intern. Med.* 2009;151(12):872–877.

13. Haddow J, Palomaki G. ACCE: A Model Process for Evaluating Data on Emerging Genetic Tests. In: Khoury M, Little J, Burke W, eds. *Human Genome Epidemiology: A Scientific Foundation for Using Genetic Information to Improve Health and Prevent Disease*. Oxford University Press; 2003:217–233.

14. Canadian Cancer Society. Colorectal Cancer Statistics at a Glance. 2012. Available at: http://www.cancer.ca/canada-wide/about cancer/cancer statistics/stats at a glance/colorectal cancer.aspx. Accessed May 30, 2012.

15. Vasen HFA, Möslein G, Alonso A, et al. Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer). *J. Med. Genet.* 2007;44(6):353–62.

16. Bakker E, Bos M, Botma A, Burema J, Croezen S, van den Donk M. *The associations between food, nutrition and physical activity and the risk of colorectal cancer and underlying mechanisms*.; 2006.

17. Little J, Sharp L. Epidemiology of colorectal cancer. In: Cassidy J, ed. *International Handbook of Colorectal Cancer*. Euromed; 2008:1–33.

18. Hippisley-Cox J, Coupland C. Identifying patients with suspected colorectal cancer in primary care: derivation and validation of an algorithm. *Br. J. Gen. Pract.* 2012;62(594):29–37.

19. Berster JM, Göke B. Type 2 diabetes mellitus as risk factor for colorectal cancer. *Arch. Physiol. Biochem.* 2008;114(1):84–98.

20. Cancer Care Ontario. Cancer Fact: More Ontarians need to be screened for colorectal cancer. *Cancer Surveill.* 2012. Available at: https://www.cancercare.on.ca/cms/one.aspx?portalId=1377&pageId=157009. Accessed December 20, 2013.

21. Imperiale T, Ransohoff D. Risk for Colorectal Cancer in Persons With a Family History of Adenomatous Polyps. *Ann. Intern. Med.* 2012;156:703–710.

22. Cancer Care Ontario. *Colon Cancer Check - 2010 Program Report*. Toronto, Canada; 2010:64.

23. Hewitson P, Glasziou P, Watson E, Towler B, Irwig L. Cochrane Systematic Review of Colorectal Cancer Screening Using the Fecal Occult Blood Test (Hemoccult): An Update. *Am. J. Gastroenterol.* 2008;103(6):1541–9.

24. Baxter NN, Goldwasser MA, Paszat LF, Saskin R, Urbach DR, Rabeneck L. Association of Colonoscopy and Death From Colorectal Cancer. *Ann. Intern. Med.* 2009;150(1):1–9.

25. Logan RF, Patnick J, Nickerson C, Coleman L, Rutter MD, von Wagner C. Outcomes of the Bowel Cancer Screening Programme (BCSP) in England after the first 1 million tests. *Gut*. 2012;61(10):1439–46.

26. D'Agostino RB, Vasan RS, Pencina MJ, et al. General Cardiovascular Risk Profile for Use in Primary Care: The Framingham Heart Study. *Circulation*. 2008;117(6):743–53.

27. Cancer Care Ontario. *Cancer Risk Factors in Ontario - Evidence Summary*.; 2013:86.

28. Cancer Care Ontario. *Focus on Colorectal Cancer in Ontario*. Toronto, Canada; 2004:28.

29. National Cancer Institute. Colorectal Cancer Risk Assessment Tool. 2010. Available at: http://www.cancer.gov/colorectalcancerrisk/. Accessed January 15, 2013.

30. Freedman AN, Slattery ML, Ballard-Barbash R, et al. Colorectal Cancer Risk Prediction Tool for White Men and Women Without Known Susceptibility. *J. Clin. Oncol.* 2009;27(5):686–693.

31. 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(5):694–8.

32. Newcomb PA, Baron J, Cotterchio M, et al. Colon Cancer Family Registry: An International Resource for Studies of the Genetic Epidemiology of Colon Cancer. *Cancer Epidemiol. Biomarkers Prev.* 2007;16(11):2331–43.

33. Cotterchio M, Mckeown-Eyssen G, Sutherland H, et al. Ontario Familial Colon Cancer Registry: Methods and First-year Response Rates. *Chronic Dis. Inj. Can.* 2000;21(2).

34. Vasen HFA, Watson P, Mecklin JP, Lynch HT. New Clinical Criteria for Hereditary Nonpolyposis Colorectal Definition (HNPCC, Lynch Syndrome) Proposed by the International Collaborative Group on HNPCC. *Gastroenterology*. 1999;116(6):1453–1456.

35. Health Canada. Body Mass Index (BMI). *Food Nutr. (Roma).* 2012. Available at: http://www.hc-sc.gc.ca/fn-an/nutrition/weights-poids/guide-ld-adult/bmi_chart_java-graph_imc_java-eng.php. Accessed June 15, 2013.

36. Health Canada. Canada's Food Guide. 2011.

37. National Cancer Institute. Pack year. *NCI Dict. Cancer Terms*. Available at: http://www.cancer.gov/dictionary?CdrID=306510. Accessed September 23, 2013.

38. Zhao J, Halfyard B, Roebothan B, et al. Tobacco Smoking and Colorectal Cancer: A Population-based Case-Control study in Newfoundland and Labrador. *Can. J. Public Heal.* 2010;101(4):281–289.

39. Canadian Pharmacists Association. Menopause. *Can. Pharm. J.* 2010;143(2).

40. Palacios S, Henderson VW, Siseles N, Tan D, Villaseca P. Age of menopause and impact of climacteric symptoms by geographical region. *Climacteric*. 2010;13(5):419–28.

41. Giovannucci E. Modifiable risk factors for colon cancer. *Gastroenterol. Clin. North Am.* 2002;31(4):925–43.

42. Sun Z, Zhu Y, Wang PP, et al. Reported Intake of Selected Micronutrients and Risk of Colorectal Cancer: Results from a Large Population-based Case-control Study in Newfoundland, Labrador and Ontario, Canada. *Anticancer Res.* 2012;32(2):687–96.

43. Cotterchio M, Manno M, Klar N, McLaughlin J, Gallinger S. Colorectal screening is associated with reduced colorectal cancer risk: a case-control study within the population-based Ontario Familial Colorectal Cancer Registry. *Cancer Causes Control*. 2005;16(7):865–75.

44. Leening M, Vedder M, Witteman J, Pencina M, Steyerberg E. Net Reclassification Improvement: Computation, Interpretation, and Controversies. *Ann. Intern. Med.* 2014;160:122–131.

45. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. 2nd ed. New York: John Wiley & Sons Inc.; 2000.

46. Hasanaj Q, Wilson BJ, Little J, Montazeri Z, Carroll JC. Family History: Impact on Coronary Heart Disease Risk Assessment beyond Guideline-Defined Factors. *Public Health Genomics*. 2013;16(5):208–14.

47. Scheuner MT, Whitworth WC, McGruder H, Yoon PW, Khoury MJ. Familial risk assessment for early-onset coronary heart disease. *Genet. Med.* 2006;8(8):525–531.

48. Scheuner MT, Setodji CM, Pankow JS, Blumenthal RS, Keeler E. General Cardiovascular Risk Profile identifies advanced coronary artery calcium and is improved by family history: the multiethnic study of atherosclerosis. *Circ. Cardiovasc. Genet.* 2010;3(1):97–105.

49. Tomeo CA, Colditz GA, Willett WC, et al. Harvard Report on Cancer Prevention. Volume 3: Prevention of colon cancer in the United States. *Cancer Causes Control*. 1999;10(3):167–80.

50. Slattery ML, Samowitz W, Hoffman M, Ma KN, Levin TR, Neuhausen S. Aspirin , NSAIDs , and Colorectal Cancer : Possible Involvement in an Insulin-Related Pathway Aspirin , NSAIDs , and Colorectal Cancer : Possible Involvement in an Insulin-Related Pathway. *Cancer Epidemiol. biomarkers Prev.* 2004;13:538–545.

51. Stolfi C, De Simone V, Pallone F, Monteleone G. Mechanisms of Action of Non-Steroidal Anti-Inflammatory Drugs (NSAIDs) and Mesalazine in the Chemoprevention of Colorectal Cancer. *Int. J. Mol. Sci.* 2013;14(9):17972–85.

52. Jang J-H, Cotterchio M, Gallinger S, Knight JA, Daftary D. Family history of hormonal cancers and colorectal cancer risk: A case-control study conducted in Ontario. *Int. J. cancer*. 2009;125(4):918–25.

53. Eppel A, Cotterchio M, Gallinger S. Allergies are associated with reduced pancreas cancer risk: A population-based case-control study in Ontario, Canada. *Int. J. Cancer*. 2007;121(10):2241–5.

54. Kohut K, Manno M, Gallinger S, Esplen MJ. Should healthcare providers have a duty to warn family members of individuals with an HNPCC-causing mutation? A survey of patients from the Ontario Familial Colon Cancer Registry. *J. Med. Genet.* 2007;44(6):404–7.

55. Sun Z, Liu L, Wang PP, et al. Association of Total Energy Intake and Macronutrient Consumption with Colorectal Cancer Risk: Results from a Large Population-based Case-control Study in Newfoundland and Labrador and Ontario, Canada. *Nutr. J.* 2012;11(1):18.

56. Levy PS, Lemeshow S. Nonresponse and Missing Data in Sample Surveys. In: *Sampling of Populations - Methods and Applications*. Fourth.; 2008:408–416.

57. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: a gentle introduction to imputation of missing values. *J. Clin. Epidemiol.* 2006;59(10):1087–91.

58. Bowen DJ, Ludman E, Press N, Vu T, Burke W. Achieving Utility with Family History. *Am. J. Prev. Med.* 2003;24(2):177–182.

59. Qureshi N, Wilson B, Santaguida P, et al. Collection and Use of Cancer Family History in Primary Care. *AHRQ - Evid. Report/Technology Assess.* 2007;(159):1–84.

60. Shields M, Gorber SC, Tremblay MS. Estimates of obesity based on self-report versus direct measures. *Heal. Reports - Stat. Canada*. 2008;19(2).

61. Neugebauer R, Ng S. Differential Recall As A Source Of Bias In Epidemiologic Research. *J. Clin. Epidemiol.* 1990;43(12):1337–41.

62. U.S. Preventive Services Task Force. Screening for Colorectal Cancer: U.S. Preventive Services Task Force Recommendation Statement. *AHRQ Publ. 08-05124-EF-3*. 2008. Available at: http://www.uspreventiveservicestaskforce.org/uspstf08/colocancer/colors.htm. Accessed January 10, 2014.

63. 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(2):216–27.

64. Church J, Lowry A, Simmang C. Practice Parameters for the Identification and Testing of Patients at Risk for Dominantly Inherited Colorectal Cancer--Supporting Documentation. *Dis. Colon Rectum*. 2001;44(10):1404–12.

65. Jellema P, van der Windt D, Bruivels D, et al. Value of symptomms and additional diagnostic test for colorectal cancer in primary care: systematic review and meta-analysis. *BMJ*. 2010;340:1–21.

66. Khoury MJ, Feero WG, Valdez R. Family History and Personal Genomics As Tools for Improving Health in an Era of Evidence-Based Medicine. *Am. J. Prev. Med.* 2010;39(2):184–8.

67. Scheuner MT, Wang SJ, Raffel LJ, Larabell SK, Rotter JI. Family history: a comprehensive genetic risk assessment method for the chronic conditions of adulthood. *Am. J. Med. Genet.* 1997;71(3):315–24. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9268102.

68. Digestive Disease Institute - Medical Informatics Group - Cleveland Clinic. Cleveland Clinic Colon Cancer Risk Assessment. 2010. Available at: http://digestive.ccf.org/scores/go. Accessed January 10, 2014.

APPENDIX A – Approval for use of ARCTIC data and data usage agreement

Application ID: C-EX-1203-02-A3

February 25, 2010

Brent Zanke, M.D., Ph.D.
Research Scientist
Department of Preventive Oncology
[redacted]
CANADA

Dear Dr. Zanke,

The amendment to your approved application, "Assessment of Risk for Colon Tumors in Canada (ARCTIC)," has been reviewed by the Colon Cancer Family Registry (Colon CFR) Steering and Advisory Committees. After careful consideration, the Advisory Committee has recommended approval of this amendment.

According to the Colon CFR records, this project is ongoing and you will continue to be contacted every year to provide information on the progress of the project and related publications.

We are pleased that your amendment has been reviewed favorably. Please do not hesitate to contact us with questions or concerns.

Very truly yours,

[redacted]

Sheri D. Schully, Ph.D.
Program Officer, Breast and Colon Cancer Family Registries
Host Susceptibility Factors Branch
Epidemiology and Genetics Research Program
Division of Cancer Control and Population Sciences
National Cancer Institute
National Institutes of Health
[redacted]

Data Usage Agreement

BREAST AND COLON CANCER FAMILY REGISTRIES DATA USE AGREEMENT

This Data Use Agreement ("Agreement") defines the responsibilities of third-party recipients and/or Institutions with respect to the use and protection of data from the Breast and Colon Cancer Family Registries ("Registries"). This data is available through Georgetown University, which serves as the National Cancer Institute's Informatics Support Center ("ISC") for the Registries. Only users whose research projects have been approved by the National Cancer Institute ("Projects") can receive data from the Registries ("Data").

All researchers seeking to use the Data and/or their employing institutions must agree to the terms of this Agreement. Researchers and/or Institutions should complete and sign in the spaces indicated below and send the signed Agreement, along with the National Cancer Institute's approval of your research project, to:

Breast and Colon Cancer Family Registries
Informatics Support Center
Georgetown University Medical Center
[redacted]

1. Information about the Researcher who is requesting the Data ("Researcher"):

[redacted] July 10, 2012
Signature Date

Brenda Joyce Wilson
Name of Researcher at Recipient Institution (printed or typed)

University of Ottawa
Institution/Organization

[redacted]
Address

[redacted] [redacted]
Telephone No. FAX No.

[redacted]
E-mail address

2. NCI Application ID and title of request

NCI Application ID: C-EX-1203-02-A3
Title request : Assessment of Risk for Colon Tumors in Canada (ARCTIC)

3. Types of Data being requested and the study population from which the Data were collected.

Data Requested:
Disease status, exposure data for various exposure as outlined in the data dictionary, family history, family history risk status for the Ontario Colorectal Registry, genomic profile of tumor

4. Proposed use and analysis plans for the Data.

Proposed Uses and Analysis:
Data will be used to conduct logistic regression analysis to determine the predictive value added by including family history in the assessment of colorectal cancer risk and impact, if any, on patient reclassification.

5. Describe the safeguards (administrative, technical, physical) that will be used by the Recipient to protect the confidentiality of the Data. Although there are alternative ways to assure security for the Data and Recipients should include the safeguards that will best meets their needs, some or all of the following safeguards must be included:

- Password protection for all files containing data (note that password protection is not regarded as sufficient protection by itself)
- Removable storage media holding the data (e.g., CDs, diskettes, zip disks, etc.) kept in a locked compartment/room when not in use
- Stored in a locked compartment/room when not in use
- No transmittal of data via e-mail, e-mail attachments, or FTP (either over the Internet, an Intranet system, or within a local area network)
- No backup copies of the data to be made
- Data stored in strongly encrypted form

You will be required to amend this section if the ISC believes that additional safeguards are necessary to protect the Data.

Safeguards to protect the confidentiality of the data:
Data will be stored in the University of Ottawa main drive, it will be accessible only to the individuals associated with the research project within the university's space.

6. If the principal Researcher (named in Section 1) is employed by an institution, list all individuals or groups of persons who will have access to or use the requested Data at the Recipient Institution, including the principal Researcher.

Identity/description of persons who will have Data access
Brenda Wilson (principal investigator)
Julian Little (co-investigator)
Leigh Jones (masters student)
Zahra Montazeri (biostatistician)
Qendresa Hasanaj (research coordinator)

7. Time period for which Data is being requested.

Estimated date of Project start: __September 2012_____

Estimated date of project completion (cannot be more than five years after Project start): September 2013_____

TERMS AND CONDITIONS

By receiving the Data described above from the ISC, the Researcher and the Recipient Institution (where applicable) agree to the following:

A. Researcher and/or Recipient Institution certify that the statements made in this Agreement (above) regarding the planned use of the Data are complete and accurate.

B. Researcher and/or Recipient Institution understand all Data will be provided on an entirely deidentified basis. Researcher and/or Recipient Institution agree that they will:

1. **Not** make any attempt to identify or contact any research subjects whose personal information is contained in the Data;

2. Use the Data only for statistical analyses and will not investigate or report the Data for specific research subjects; and

3. Notify the ISC immediately if the identity of any research subject is discovered inadvertently.

C. Researcher and/or Recipient Institution will not use the Data for purposes other than described in this Agreement.

D. Researcher and/or Recipient Institution will establish and maintain the administrative, technical, and physical safeguards, described in Section 5 above to protect the confidentiality of the data and to prevent unauthorized use or access to the Data.

E. Research and/or Recipient Institution will not disclose the Data to anyone outside Recipient Institution and will only disclose the Data to those individuals or groups at Recipient Institutions that are listed in Section 6 above. Within the Recipient Institution, access to the Data shall be limited to the minimum number of individuals necessary to achieve the purpose stated in the Agreement. Researcher will ensure that all individuals within Recipient Institution that have access to the Data are aware of and will abide by the terms of this Agreement. If Researcher is not employed by an institution, Researcher will not share the Data with anyone else.

F. No findings or information derived from the Data may be released if such findings contain any combination of data elements that might allow for identification or the deduction of a study participant's identity.

G. Researcher and/or Recipient Institution agree to subject any findings or manuscripts proposed for public release (e.g., abstracts, presentations, publications) to a stringent review to assure that data confidentiality is maintained and that individual study participants cannot be identified.

H. Researcher and/or Recipient Institution understand that data is experimental in nature and is provided on an "as is" basis, without any express or implied warranties, including but not limited to fitness for a particular purpose.

I. Researcher or Recipient Institution will report immediately to the ISC any use or disclosure of the Data other than as permitted by this Agreement, and will take all reasonable steps to mitigate the effects of such improper use or disclosure, cooperating with all reasonable requests by the Provider towards that end.

J. This Agreement will become effective on the date the ISC sends the Data to Recipient and will remain in effect for five years or until six months after the completion of the Project, whichever is earlier. Either the Recipient or the ISC may terminate this Agreement upon thirty days written notice.

K. Upon expiration or termination of this Agreement, Researcher and/or Recipient Institution will destroy all copies of Data or portions thereof in its possession that were received from the Provider or created (or had others create) using Data received from the Provider ("Derivative Data Sets") ***unless Recipient Institution's policies dictate otherwise***. Researcher and/or Recipient Institution will notify the ISC of the destruction of the Data and the Derivative Data Sets ***or will indicate why they cannot destroy the data***. In the event that Recipient Institution's policies do not permit destruction of the Data and Derivative Data

to the provision of this Agreement directly involved in the controversy in which such judgment shall have been rendered.

Signature of Recipient Institution (if Researcher is employed by an institution/organization):

The undersigned individual hereby attests that he or she is authorized to legally bind the Recipient Institution to the terms of this Agreement and agrees to all the terms specified herein.

[redacted] JUL 16 2012

Signature Date

Joanne R. Lauzon
Directrice adjointe, VRTT
Assistant Director, TTBE

Name of Official from Recipient Institution (printed or typed)

Institution/Organization

[redacted]

Telephone No. [redacted] E-mail Address

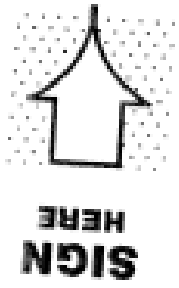
Signature of Researcher:

[redacted] July 10, 2012

Signature Date

BCENOA JOYCE WILSON

Name of Researcher (printed or typed)



Sets, Recipient Institution will continue to maintain the Data and Derivative Data Sets consistent with all of the restrictions contained in this Agreement.

L. If the Researcher leaves the employ of the Recipient Institution, Recipient Institution will notify the ISC in writing at least fourteen days before the Researcher leaves. Researcher is not permitted to take or use the Data or any Derivative Data Sets following the end of his or her employment with the Recipient Institution. ***The Researcher must arrange a new Data Use Agreement between the ISC and his/her new employer in order to continue using the Data or any Derivative Data Sets.*** Within fourteen days after the Researcher leaves the Recipient Institution, Recipient Institution agrees to destroy the Data and all Derivative Data Sets ***unless institutional policies dictate otherwise.*** Researcher and/or Recipient Institution will notify the ISC of the destruction of the Data and the Derivative Data Sets ***or will indicate why they cannot destroy the data***. In the event that Recipient Institution's policies do not permit destruction of the Data and Derivative Data Sets, Recipient Institution will continue to maintain the Data and Derivative Data Sets consistent with all of the restrictions contained in this Agreement.

M. In the event that the ISC determines or has a reasonable belief that Recipient Institution or Researcher has violated any terms of this Agreement, the ISC may take any of the following actions:

1. Revoke the existing Agreement and demand the destruction of the Data and all Derivative Data Sets.
2. Deny Researcher and/or Recipient Institution future access to data from the Registries.
3. Report the violation to Researcher's Recipient Institution (if applicable) for action pursuant to the institution's policies.
4. If the confidentiality of human subjects has been violated, the case may be reported to the U.S. Department of Health and Human Services' Office for Human Research Protection ("OHRP") for investigation and possible action against a Researcher or Recipient Institution.
5. The ISC or NCI may seek an injunction or damages against Researcher and/or Recipient Institution in a court of competent jurisdiction.

N. Researcher and Recipient Institution will comply with all applicable laws in their use of the Data. This Agreement shall be construed in a manner that supports compliance by Recipient and Provider with all applicable requirements of HIPAA (Health Insurance Portability and Accountability Act), the Privacy Rule, and the Privacy Act of 1974.

O. This Agreement shall be governed by the laws of the State of New York, USA. This Agreement contains the entire agreement with the ISC concerning the subject matter hereof. No modifications of this Agreement or waiver of the terms and conditions hereof will be binding upon, unless approved in writing by the ISC. The Recipient Institution and/or the Researcher may not assign this Agreement without the ISC's written consent. If any provision of this Agreement shall, for any reason, be adjudged by any court of competent jurisdiction to be invalid or unenforceable, such judgment shall not affect, impair or invalidate the remainder of this Agreement but shall be confined in its operation

APPENDIX B – OHREB Project Ethics Approval

Ottawa Hospital Research Ethics Boards / Conseils d'éthique en recherches
725 Parkdale Avenue, Box 411, Ottawa, Ontario K1Y 4E9 613-798-5555 ext. 14902 Fax: 613-761-4311
http://www.ohri.ca/ohreb

December 21, 2012

Dr. Brenda Wilson
University of Ottawa
Epidemiology & Community Medicine
[redacted]

Dear Dr. Wilson:

Re: Protocol # 20120950-01H **The clinical validity of family history in risk stratification of colorectal cancer**

Protocol approval valid until - December 20, 2013

Thank you for the email from Leigh Jonah dated December 20, 2012. I am pleased to inform you that this protocol underwent expedited review by the Ottawa Hospital Research Ethics Board (OHREB) and is approved. No changes, amendments or addenda may be made to the protocol without the OHREB's review and approval.

Approval is for the following:
- Electronic OHREB Application
- Thesis Proposal, dated August 24, 2012

If the study is to continue beyond the expiry date noted above, a Renewal Form should be submitted to the OHREB approximately six weeks prior to the current expiry date. If the study has been completed by this date, a Termination Report should be submitted.

The Ottawa Hospital Research Ethics Board is constituted in accordance with, and operates in compliance with the requirements of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans; Health Canada Good Clinical Practice: Consolidated Guideline; Part C Division 5 of the Food and Drug Regulations of Health Canada; and the provisions of the Ontario Health Information Protection Act 2004 and its applicable Regulations.

Yours sincerely,
[redacted]
Raphael Saginur, M.D.
Chairman
Ottawa Hospital Research Ethics Board

RS/ll

Research Institute
Institut de recherche

uOttawa

UNIVERSITY OF OTTAWA HEART INSTITUTE
INSTITUT DE CARDIOLOGIE DE L'UNIVERSITÉ D'OTTAWA

***Ottawa Health Science Network Research Ethics Board/* Réseau des sciences de la santé d'Ottawa Conseil d'éthique de la recherche**

Civic Box 411 725 Parkdale Avenue, Ottawa, Ontario K1Y 4E9 613-798-5555 ext. 14902 Fax : 613-761-4311
http://www.ohri.ca/ohsn-reb

November 20, 2013

Dr. Brenda Wilson
University of Ottawa
Epidemiology & Community Medicine
[redacted]

Dear Dr. Wilson:

RE: Protocol# - 20120950-01H **The clinical validity of family history in risk stratification of colorectal cancer**

Renewal Expiry Date - December 20, 2014

Thank you for the email dated October 21, 2013 from Brenda Wilson. I am pleased to inform you that your Annual Renewal Request (listed above) was reviewed by the Ottawa Health Science Network Research Ethics Board (OHSN-REB) and is approved. No changes, amendments or addenda may be made in the protocol without the OHSN-REB's review and approval.

Renewal is valid for a period of one year. Approximately one month prior to that time, a single renewal form should be sent to the REB office.

Also acknowledged is the projected date of study completion to be extended to December 20, 2014.

The Ottawa Health Science Network Research Ethics Board (OHSN-REB) was created by the merger of both the Ottawa Hospital Research Ethics Board (OHREB) and the Human Research Ethics Board (HREB) for meetings held at the University of Ottawa Heart Institute.

OHSN-REB complies with the membership requirements and operates in compliance with the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans; the International Conference on Harmonization - Good Clinical Practice: Consolidated Guideline; and the provisions of the Personal Health Information Protection Act 2004.

The Tri-Council Policy Statement requires a greater involvement of the OHSN-REB in studies over the course of their execution. As well, you must inform the Board of adverse events encountered during the study, here or elsewhere, or of significant new information which becomes available after the Board review, either of which may impinge on the ethics of continuing the study. The OHSN-REB will review the new information to determine if the protocol should be modified, discontinued, or should continue as originally approved.

Yours sincerely,

[redacted]

Raphael Sa̋ginur, M.D.
Chairperson
Ottawa Health Science Network Research Ethics Board
/jm

Ontario
Familial Colon Cancer Registry

Family History Questionnaire

Should you wish to talk to someone about this questionnaire,
you may call [redacted]

If you don't know an answer, please write "Don't know" or "DK" in the space for the answer.

If you are not sure of a date, please make the best guess you can, and put a question mark beside it.

If there is not enough space to list all your relatives, please write on the inside of the front cover, or the outside of the back cover.

If you were adopted , please write "adopted" on this page, answer Section 1 about yourself on this page, and answer any questions you can about your biological (blood) relatives, including any children you may have.

Section 1: Yourself

1.1 **Name** _____ **Date of birth** day/month/year ___ / ___ / ___

Maiden name _____ **Any other last names** _____

Telephone (_____) _____ _____ (home)

(_____) _____ _____ (work)

1.2 Have you ever had cancer diagnosed before your most recently diagnosed colon cancer?

- ○ No
- ○ Yes

→ **Type of cancer or tumour** **Date of diagnosis** (day/month/year)

_____ ___ / ___ / ___

_____ ___ / ___ / ___

Please list your parents, children, brothers and sisters on the following pages, referring to records or asking other family members for information when you need to. The last page asks whether any of your other relatives have had cancer.

If you don't know an answer, please write "Don't know" or "DK" in the space for the answer.
If you are not sure of a date, please make your best guess and put a question mark beside it.

Section 2: Your Mother

2.1 **Full name**
(First/middle/last)

_____

Age _____

Date of birth
(day/month/year)
___ / ___ / ___

Maiden name

_____

Any previous married name(s)

_____

2.2 Is your mother alive?

- ○ No
- ○ Don't know → *Please go to question 2.3.*
- ○ Yes → *Please go to question 2.3.*

→ **Cause of death**

_____

Age at death _____

Date of death
(day/month/year)
___ / ___ / ___

2.3 Has your mother had any cancers or tumours?

Include any leukemia or lymphoma. If there was more than one cancer or tumour, please list them all.
If the cancer started in one place and spread, please indicate where it started.

- ○ No → *Please go to Section 3, page 3.*
- ○ Don't know → *Please go to Section 3, page 3.*
- ○ Yes

→ **Type of cancer or tumour**

_____ **Age at diagnosis** _____ **Date of diagnosis** (day/month/year) ___ / ___ / ___

_____ _____ ___ / ___ / ___

Section 3: Your Father

3.1 **Full name**
(First/middle/last)

Age

Date of birth
(day/month/year)

___ / ___ / ___

3.2 Is your father alive?

- ○ No
- ○ Don't know → *Please go to question 3.3.*
- ○ Yes → *Please go to question 3.3.*

Cause of death

Age at death

Date of death
(day/month/year)

___ / ___ / ___

3.3 Has your father had any cancers or tumours?

Include any leukemia or lymphoma. If there was more than one cancer or tumour, please list them all. If the cancer started in one place and spread, please indicate where it started.

- ○ No → *Please go to Section 4, pages 4 and 5.*
- ○ Don't know → *Please go to Section 4, pages 4 and 5.*
- ○ Yes

Type of cancer or tumour

Age at diagnosis

Date of diagnosis
(day/month/year)

___ / ___ / ___

___ / ___ / ___

If you don't know an answer, please write "Don't know" or "DK" in the space for the answer.
If you are not sure of a date, please make your best guess and put a question mark beside it.

Section 4: Your Children

4.1 How many children have you had? _____ *If none, please go to Section 5, pages 6 and 7.*

- *Please list them all, living and deceased (write on the inside of the front cover or the outside of the back cover if necessary).*
- *If any have changed their last names, by marriage or otherwise, please list the last names they use now.*
- *If you adopted any of your children, please write "adopted" beside their names.*
- *If some of your children had different fathers, please note that fact beside their names.*

Full name (First/middle/last)	**Sex** (Circle)	**Age**	**Date of birth** (day/month/year)
_____	M F	_____	___ / ___ / ___
_____	M F	_____	___ / ___ / ___
_____	M F	_____	___ / ___ / ___
_____	M F	_____	___ / ___ / ___
_____	M F	_____	___ / ___ / ___
_____	M F	_____	___ / ___ / ___
_____	M F	_____	___ / ___ / ___
_____	M F	_____	___ / ___ / ___
_____	M F	_____	___ / ___ / ___
_____	M F	_____	___ / ___ / ___

4.2 Are all your children alive?

- ○ No
- ○ Don't know → *Please go to question 4.3.*
- ○ Yes → *Please go to question 4.3.*

Names of any children who have died	**Cause of death**	**Age at death**	**Date of death** (day/month/year)
			___ / ___ / ___
			___ / ___ / ___
			___ / ___ / ___
			___ / ___ / ___

4.3 Have any of your children had any cancers or tumours?

Include any leukemia or lymphoma. If there was more than one cancer or tumour, please list them all. If the cancer started in one place and spread, please indicate where it started.

- ○ No → *Please go to Section 5, pages 6 and 7.*
- ○ Don't know → *Please go to Section 5, pages 6 and 7.*
- ○ Yes

Names of any children who have had cancer	**Type of cancer or tumour**	**Age at diagnosis**	**Date of diagnosis** (day/month/year)
			___ / ___ / ___
			___ / ___ / ___
			___ / ___ / ___
			___ / ___ / ___

If you don't know an answer, please write "Don't know" or "DK" in the space for the answer.
If you are not sure of a date, please make your best guess and put a question mark beside it.

Section 5: Your Brothers and Sisters

5.1 How many brothers and sisters do you have (living and deceased)? _____ *If none, please go to Section 6, pages 8 and 9.*

- *Please list them all, living and deceased (write on the inside of the front cover or the outside of the back cover if necessary).*
- *If any have changed their last names, by marriage or otherwise, please list the last names they use now.*
- *If any of your brothers or sisters were adopted please write "adopted" beside their names.*
- *If some are your half-brothers or half-sisters, please write "half-brother" or "half-sister" beside their names, and whether they had the same mother or same father as yourself.*
- *Please do not list any step-brothers or step-sisters (children from your step-mother's or step-father's previous marriage).*

Full name (First/middle/last)	**Sex** (Circle)	**Age**	**Date of birth** (day/month/year)
_____	M F	_____	___ / ___ / ___
_____	M F	_____	___ / ___ / ___
_____	M F	_____	___ / ___ / ___
_____	M F	_____	___ / ___ / ___
_____	M F	_____	___ / ___ / ___
_____	M F	_____	___ / ___ / ___
_____	M F	_____	___ / ___ / ___
_____	M F	_____	___ / ___ / ___
_____	M F	_____	___ / ___ / ___
_____	M F	_____	___ / ___ / ___

5.2 Are all your brothers and sisters alive?

- ○ No
- ○ Don't know → *Please go to question 5.3.*
- ○ Yes → *Please go to question 5.3.*

Names of any brothers and sisters who have died	**Cause of death**	**Age at death**	**Date of death** (day/month/year)
			___ / ___ / ___
			___ / ___ / ___
			___ / ___ / ___
			___ / ___ / ___
			___ / ___ / ___

5.3 Has any of your brothers or sisters had any cancers or tumours?

Include any leukemia or lymphoma. If anyone had more than one cancer or tumour, please list them all. If the cancer started in one place and spread, please indicate where it started.

- ○ No → *Please go to Section 6, pages 8 and 9.*
- ○ Don't know → *Please go to Section 6, pages 8 and 9.*
- ○ Yes

Names of any brothers or sisters who have had cancer	**Type of cancer or tumour**	**Age at diagnosis**	**Date of diagnosis** (day/month/year)
			___ / ___ / ___
			___ / ___ / ___
			___ / ___ / ___
			___ / ___ / ___

If you don't know an answer, please write "Don't know" or "DK" in the space for the answer.
If you are not sure of a date, please make your best guess and put a question mark beside it.

Section 6: Your Other Relatives Who Have Had Cancer

6.1 Have any of your other relatives (for example, grandparents, aunts and uncles, nieces and nephews, first or second cousins, grandchildren) had any cancers or tumours?

- *Please write which side of the family any relative was on (for example, "mother's mother", "father's sister").*
- *Include any leukemia or lymphoma. If anyone had more than one cancer or tumour, please list them all. If the cancer started in one place and spread, please indicate where it started.*
- *If any of these relatives were half-brothers or half-sisters of your parents, please write, for example, "mother's half-brother".*
- *If any relatives you list were adopted into the family, please write "adopted" beside the line where you list them.*

○ No

○ Don't know

○ Yes

Name	**Relationship to you** (for example, mother's father, or cousin on father's side)	**Type of cancer or tumour**	**Age at diagnosis**	**Is this relative alive?**

We recognize that the following question may be sensitive for some people. If you do not want to answer, it would be helpful to us if you would write "prefer not to answer" beside it.

Section 7: Your Ethnic or Racial Background

7.1 What is your ethnic or racial background? *(Please tick as many as apply.)*

O Black	O Middle Eastern (e.g. Egyptian, Lebanese, Iranian)	O Other
O White	O Filipino	*Please specify* ____________
O Native (e.g. Indian, Inuit)	O East Asian (e.g. Chinese, Korean, Vietnamese)	O Don't know
O Jewish	O South Asian (e.g. East Indian, Pakistani)	

9

Source: Cancer Family Registries – Informatics Support Centre.
http://www.coloncfr.org/questionnaires/Ontario/Ontario-FamilyHistory.pdf

Ontario
Familial Colon Cancer Registry

Personal History Questionnaire

This questionnaire is about factors that may relate to a person's risk of developing cancer. Although it is important to have complete data for scientific reasons and we encourage you to answer all questions, if you come to a question that you do not want to answer, please write "prefer not to answer" beside it and then continue to answer the remaining questions.

Should you wish to talk to someone about this questionnaire, you may call (416) 946-4409 or 1-800-832-5949.

Please write in your answers where space is provided, or place tick marks in circles ☑

What date are you filling out this questionnaire? ___ / ___ / ___
day month year

Identifying Information

1. Are you male or female?
   - ○ male
   - ○ female

2. What is your age?
   - ___ ___ years
   - ○ don't know

3. What is your date of birth?
   - day ___ ___
   - month ___ ___
   - year ___ ___ ___ ___
   - ○ don't know day
   - ○ don't know month
   - ○ don't know year

4. Are you a twin or triplet?
   - ○ yes, a twin
   - ○ yes, other multiple (triplet, quadruplet, etc.): _____
     please specify
   - ○ no
   - ○ don't know

   → *If yes, please read the following statement and answer the question.*

   Non-identical twins are no more alike than ordinary brothers and sisters. Genetically identical twins, on the other hand, look so much alike (that is, they have a strong resemblance to each other in height, colouring, features of the face, etc.) that people often mistake one for the other, especially during their childhood.

   Do you have a genetically identical twin or triplet?
   - ○ yes
   - ○ no
   - ○ don't know

5. What is your marital status?
   - ○ currently married or living as married
   - ○ separated
   - ○ divorced
   - ○ widowed
   - ○ single or never married
   - ○ don't know

Bowel Screening and Health

6. Have you ever had a **test for blood in your stool, called a smear test or a hemoccult**? This test is frequently done as part of a routine physical examination, or it can be done at home.

- O yes
- O no → *Please go to #7*
- O don't know → *Please go to #7*

6a. When did you **first** have this test?

age when **first** tested ___ ___

or

year of **first** test ___ ___ ___ ___

- O don't know

6b. What were the reasons for your **first** test? *Please tick all that apply.*

- O to investigate a new problem
- O family history of colorectal cancer
- O routine/yearly examination or check-up
- O follow-up of a previous problem
- O other: _____ *please specify*
- O don't know

6c. How many times have you had a hemoccult test?

___ ___ number of hemoccult tests

- O don't know

6d. If you have had a hemoccult test more than once, when did you **last** have this test?

age when **last** tested ___ ___

or

year of **last** test ___ ___ ___ ___

- O don't know

7. Have you ever had a **sigmoidoscopy**? Sigmoidoscopy involves looking inside the lower bowel and rectum with a lighted instrument. This examination is usually done in a doctor's office without anesthesia.

- O yes
- O no → *Please go to #8*
- O don't know → *Please go to #8*

7a. When did you **first** have this test?

age when **first** tested ___ ___

or

year of **first** test ___ ___ ___ ___

- O don't know

7b. What were the reasons for your **first** sigmoidoscopy? *Please tick all that apply.*

- O to investigate a new problem
- O family history of colorectal cancer
- O routine/yearly examination or check-up
- O follow-up of a previous problem
- O other: _____ *please specify*
- O don't know

7c. How many times have you had a sigmoidoscopy?

___ ___ number of sigmoidoscopies

- O don't know

7d. If you have had a sigmoidoscopy more than once, when did you **last** have this test?

age when **last** tested ___ ___

or

year of **last** test ___ ___ ___ ___

- O don't know

8. Have you ever had a **colonoscopy**? Colonoscopy is an examination of the entire large bowel using a long flexible instrument. This examination is usually done under sedation.

- O yes
- O no → *Please go to #9*
- O don't know → *Please go to #9*

8a. When did you **first** have this test?

age when **first** tested ___ ___

or

year of **first** test ___ ___ ___ ___

- O don't know

8b. What were the reasons for your **first** colonoscopy? *Please tick all that apply.*

- O to investigate a new problem
- O family history of colorectal cancer
- O routine/yearly examination or check-up
- O follow-up of a previous problem
- O other: _____ *please specify*
- O don't know

8c. How many times have you had a colonoscopy?

___ ___ number of colonoscopies

- O don't know

8d. If you have had a colonoscopy more than once, when did you **last** have this test?

age when **last** tested ___ ___

or

year of **last** test ___ ___ ___ ___

- O don't know

9. Has a doctor ever told you that you had **polyps** in your large bowel or colon or rectum? Polyps are growths in the lining of the colon which vary in size from a tiny dot to several inches.

- O yes
- O no → *Please go to #10*
- O don't know → *Please go to #10*

9a. When did your doctor **first** tell you that you had polyps?

age at **first** diagnosis ___ ___

or

year of **first** diagnosis ___ ___ ___ ___

- O don't know

9b. Have you been told more than once that you had polyps?

- O yes
- O no
- O don't know

9c. When did your doctor **last** tell you that you had polyps?

age at **last** diagnosis ___ ___

or

year of **last** diagnosis ___ ___ ___ ___

- O don't know

9d. Do you know what kind of polyps they were? *Please include all the separate times you were told you had polyps. Please tick all that apply.*

- O benign
- O adenomatous (pre-cancerous)
- O hyperplastic
- O other: _____ *please specify*
- O don't know

9e. Did you have the polyps removed (by a procedure called a polypectomy)? (This can be done during a sigmoidoscopy or colonoscopy.)

- O yes
- O no → *Please go to #10*
- O don't know → *Please go to #10*

9f. When did you **first** have polyps removed?

age at **first** polypectomy ___ ___

or

year of **first** polypectomy ___ ___ ___ ___

- O don't know

9g. Have you had polyps removed more than once?

- O yes
- O no
- O don't know

9h. If you have had polyps removed more than once, when did you **last** have polyps removed?

age at **last** polypectomy ___ ___

or

year of **last** polypectomy ___ ___ ___ ___

- O don't know

10. Has a doctor ever told you that you had **familial adenomatous polyposis, known also as FAP**? This is a condition, sometimes occurring in families, in which numerous polyps line the inside of the large bowel or colon.

- O yes
- O no → *Please go to #11*
- O don't know → *Please go to #11*

10a. When did your doctor **first** tell you that you had FAP?

age at diagnosis ___ ___

or

year of diagnosis ___ ___ ___ ___

- O don't know

11. Has a doctor ever told you that you had **Crohn's disease**? This is where you have an inflammation that extends into the deeper layers of the intestinal wall. It may also affect other parts of the digestive tract, including the mouth, esophagus, stomach, and small intestine.

- O yes
- O no → *Please go to #12*
- O don't know → *Please go to #12*

11a. When did your doctor **first** tell you that you had Crohn's disease?

age at diagnosis ___ ___

or

year of diagnosis ___ ___ ___ ___

- O don't know

12. Has a doctor ever told you that you had **ulcerative colitis**? This is an inflammation and ulceration of the lining of the bowel (colon) and rectum. It is not a stomach ulcer.

- O yes
- O no → *Please go to #13*
- O don't know → *Please go to #13*

12a. When did your doctor **first** tell you that you had ulcerative colitis?

age at diagnosis ___ ___

or

year of diagnosis ___ ___ ___ ___

- O don't know

13. Has a doctor ever told you that you had **irritable bowel syndrome**? This is a disorder of the bowels leading to cramping, gassiness, bloating and alternating diarrhoea and constipation. It is sometimes called IBS, or spastic colon.

- O yes
- O no → *Please go to #14*
- O don't know → *Please go to #14*

13a. When did your doctor **first** tell you that you had irritable bowel syndrome?

age at diagnosis ___ ___

or

year of diagnosis ___ ___ ___ ___

○ don't know

14. Has a doctor ever told you that you had **diverticular disease**? This may also be called diverticulosis or diverticulitis. It's a condition in which the bowel may become infected, and can lead to pain and chronic problems with bowel habits.

○ yes

○ no → *Please go to #15*

○ don't know → *Please go to #15*

14a. When did your doctor **first** tell you that you had diverticular disease?

age at diagnosis ___ ___

or

year of diagnosis ___ ___ ___ ___

○ don't know

15. Have you ever had any of your **large bowel or colon** removed?

○ yes

○ no → *Please go to #16*

○ don't know → *Please go to #16*

→ Was it completely removed, or was only part of it removed?

○ completely removed

○ partly removed

○ don't know

15a. When did you **first** have any of your bowel or colon removed?

age at **first** operation ___ ___

or

year of **first** operation ___ ___ ___ ___

○ don't know

15b. Have you had more than one surgery to remove your bowel or colon?

○ yes

○ no → *Please go to #16*

○ don't know → *Please go to #16*

15c. When did you **last** have all or part of your bowel or colon removed?

age at **last** operation ___ ___

or

year of **last** operation ___ ___ ___ ___

○ don't know

16. Have you had your **gallbladder** removed?

○ yes

○ no → *Please go to #17*

○ don't know → *Please go to #17*

16a. When did you have your gallbladder removed?

age at operation ___ ___

or

year of operation ___ ___ ___ ___

○ don't know

17. Has a doctor ever told you that you had **diabetes**, also known as **diabetes mellitus**? *Please do not include diabetes which you had **only** during pregnancy.*

○ yes

○ no → *Please go to #18*

○ don't know → *Please go to #18*

17a. When did your doctor **first** tell you that you had diabetes?

age at diagnosis ___ ___

or

year of diagnosis ___ ___ ___ ___

○ don't know

17b. Did you ever take medication to control your diabetes?

- ○ yes
- ○ no → *Please go to #18*
- ○ don't know → *Please go to #18*

17c. What type of medication did you use, pills or insulin injections?

- ○ pills
- ○ insulin injections
- ○ both
- ○ don't know → *Please go to #18*

17d. How often did you usually take it?
Please choose the most appropriate category.

	Pills	**Insulin**
times per day *or*	___ ___	___ ___
times per week *or*	___ ___	___ ___
times per month *or*	___ ___	___ ___
times per year	___ ___	___ ___
don't know	○	○

17e. **About two years** ago, were you taking it?

	Pills	**Insulin**
yes	○	○
no	○	○
don't know	○	○

17f. How long, in total, have you taken this medication?

	Pills	**Insulin**
number of months *or*	___ ___	___ ___
number of years	___ ___	___ ___
don't know	○	○

18. Has a doctor ever told you that you had **high cholesterol**? *If your doctor told you it was borderline, please tick no.*

- ○ yes
- ○ no → *Please go to #19*
- ○ don't know → *Please go to #19*

18a. When did your doctor **first** tell you that you had high cholesterol?

age at diagnosis ___ ___
or
year of diagnosis ___ ___ ___ ___

- ○ don't know

18b. Did you ever take medication to control your high cholesterol?

- ○ yes
- ○ no → *Please go to #19*
- ○ don't know → *Please go to #19*

18c. How often did you usually take it?
Please choose the most appropriate category.

- ___ ___ times per day *or*
- ___ ___ times per week *or*
- ___ ___ times per month *or*
- ___ ___ times per year
- ○ don't know

18d. **About two years** ago, were you taking it?

- ○ yes
- ○ no
- ○ don't know

18e. How long, in total, have you taken this medication?

- ___ ___ number of months *or*
- ___ ___ number of years
- ○ don't know

19. Has a doctor ever told you that you had **high levels of fat (other than cholesterol) in your blood,** also called **high triglycerides**? *If your doctor told you it was borderline, please tick no.*

- O yes
- O no → *Please go to #20*
- O don't know → *Please go to #20*

19a. When did your doctor **first** tell you that you had high triglycerides?

age at diagnosis ___ ___

or

year of diagnosis ___ ___ ___ ___

- O don't know

19b. Did you ever take medication to control the high levels of fat in your blood?

- O yes
- O no → *Please go to #20*
- O don't know → *Please go to #20*

19c. How often did you usually take it? *Please choose the most appropriate category.*

___ ___ times per day *or*

___ ___ times per week *or*

___ ___ times per month *or*

___ ___ times per year

- O don't know

19d. **About two years** ago, were you taking it?

- O yes
- O no
- O don't know

19e. How long, in total, have you taken this medication?

___ ___ number of months *or*

___ ___ number of years

- O don't know

20. Has a doctor ever told you that you had any type of **cancer**?

- O yes
- O no → *Please go to #24*
- O don't know → *Please go to #24*

20a. What type of cancer was it?

_____ cancer

20b. When did your doctor **first** tell you that you had this type of cancer?

age at diagnosis ___ ___

or

year of diagnosis ___ ___ ___ ___

- O don't know

20c. Were you treated with radiation therapy (radiotherapy) for this cancer?

- O yes
- O no
- O don't know

21. Has a doctor ever told you that you had any other **cancer**?

- O yes
- O no → *Please go to #24*
- O don't know → *Please go to #24*

21a. What type of cancer was it?

_____ cancer

21b. When did your doctor **first** tell you that you had this type of cancer?

age at diagnosis ___ ___

or

year of diagnosis ___ ___ ___ ___

- O don't know

21c. Were you treated with radiation therapy (radiotherapy) for this cancer?

- O yes
- O no
- O don't know

22. Has a doctor ever told you that you had any other **cancer**?

- O yes
- O no → *Please go to #24*
- O don't know → *Please go to #24*

22a. What type of cancer was it?

_____ cancer

22b. When did your doctor **first** tell you that you had this type of cancer?

age at diagnosis __ __

or

year of diagnosis __ __ __ __

- O don't know

22c. Were you treated with radiation therapy (radiotherapy) for this cancer?

- O yes
- O no
- O don't know

23. Has a doctor ever told you that you had any other **cancer**?

- O yes
- O no → *Please go to #24*
- O don't know → *Please go to #24*

23a. What type of cancer was it?

_____ cancer

23b. When did your doctor **first** tell you that you had this type of cancer?

age at diagnosis __ __

or

year of diagnosis __ __ __ __

- O don't know

23c. Were you treated with radiation therapy (radiotherapy) for this cancer?

- O yes
- O no
- O don't know

Medications

Have you ever taken any of the following medications regularly (at least twice a week for more than a month)?

24. **Aspirin (such as Anacin, Bufferin, Bayer, Excedrin, Ecotrin)**

- O yes
- O no → *Please go to #25*
- O don't know → *Please go to #25*

24a. How often did you usually take it when you were taking it regularly (that is, at least twice a week for more than a month)?
Please choose one of the following.

__ __ times per day *or*

__ __ times per week

- O don't know

24b. **About two years** ago, were you taking it regularly?

- O yes
- O no
- O don't know

24c. How long, in total, have you taken this medication regularly? *If you started and stopped and then started again, please count only the time you were taking this medication.*

__ __ number of months *or*

__ __ number of years

- O don't know

Have you ever taken any of the following medications regularly (at least twice a week for more than a month)? *(continued)*

25. **Acetaminophen (such as Tylenol, Anacin-3, Panadol)**

- ○ yes
- ○ no → *Please go to #26*
- ○ don't know → *Please go to #26*

25a. How often did you usually take it when you were taking it regularly (that is, at least twice a week for more than a month)? *Please choose one of the following.*

- ___ ___ times per day *or*
- ___ ___ times per week
- ○ don't know

25b. **About two years** ago, were you taking it regularly?

- ○ yes
- ○ no
- ○ don't know

25c. How long, in total, have you taken this medication regularly? *If you started and stopped and then started again, please count only the time you were taking this medication.*

- ___ ___ number of months *or*
- ___ ___ number of years
- ○ don't know

26. **Ibuprofen-based medications (such as Advil, Motrin, Nuprin, Medipren, Indocid, Naprosyn, NSAIDS (NSAIDS are non-steroidal anti-inflammatory drugs)**

- ○ yes
- ○ no → *Please go to #27*
- ○ don't know → *Please go to #27*

26a. How often did you usually take it when you were taking it regularly (that is, at least twice a week for more than a month)? *Please choose one of the following.*

- ___ ___ times per day *or*
- ___ ___ times per week
- ○ don't know

26b. **About two years** ago, were you taking it regularly?

- ○ yes
- ○ no
- ○ don't know

26c. How long, in total, have you taken this medication regularly? *If you started and stopped and then started again, please count only the time you were taking this medication.*

- ___ ___ number of months *or*
- ___ ___ number of years
- ○ don't know

Have you ever taken any of the following medications regularly (at least twice a week for more than a month)? *(continued)*

27. **Bulk-forming laxatives (such as Metamucil, Citrucel, FiberCon, Serutan, psyllium)**

- ○ yes
- ○ no → *Please go to #28*
- ○ don't know → *Please go to #28*

27a. How often did you usually take it when you were taking it regularly (that is, at least twice a week for more than a month)? *Please choose one of the following.*

- ___ ___ times per day *or*
- ___ ___ times per week
- ○ don't know

27b. **About two years** ago, were you taking it regularly?

- ○ yes
- ○ no
- ○ don't know

27c. How long, in total, have you taken this medication regularly? *If you started and stopped and then started again, please count only the time you were taking this medication.*

- ___ ___ number of months *or*
- ___ ___ number of years
- ○ don't know

28. **Other laxatives (such as Ex-Lax, Correctol, Dulcolax, Senokot, Colace, castor oil, cod liver oil, mineral oil, milk of magnesia, lactulose, Epsom salts)**

- ○ yes
- ○ no → *Please go to #29*
- ○ don't know → *Please go to #29*

28a. How often did you usually take it when you were taking it regularly (that is, at least twice a week for more than a month)? *Please choose one of the following.*

- ___ ___ times per day *or*
- ___ ___ times per week
- ○ don't know

28b. **About two years** ago, were you taking it regularly?

- ○ yes
- ○ no
- ○ don't know

28c. How long, in total, have you taken this medication regularly? *If you started and stopped and then started again, please count only the time you were taking this medication.*

- ___ ___ number of months *or*
- ___ ___ number of years
- ○ don't know

Have you ever taken any of the following medications regularly (at least twice a week for more than a month)? *(continued)*

29. **Multivitamin supplements (such as One-A-Day, Theragram, Centrum, Unicap)** (not individual vitamins)

- ○ yes
- ○ no → *Please go to #30*
- ○ don't know → *Please go to #30*

29a. How often did you usually take it when you were taking it regularly (that is, at least twice a week for more than a month)? *Please choose one of the following.*

- ___ ___ times per day *or*
- ___ ___ times per week
- ○ don't know

29b. **About two years** ago, were you taking it regularly?

- ○ yes
- ○ no
- ○ don't know

29c. How long, in total, have you taken this medication regularly? *If you started and stopped and then started again, please count only the time you were taking this medication.*

- ___ ___ number of months *or*
- ___ ___ number of years
- ○ don't know

30. **Folic acid or folate pills or tablets**

- ○ yes
- ○ no → *Please go to #31*
- ○ don't know → *Please go to #31*

30a. How often did you usually take it when you were taking it regularly (that is, at least twice a week for more than a month)? *Please choose one of the following.*

- ___ ___ times per day *or*
- ___ ___ times per week
- ○ don't know

30b. **About two years** ago, were you taking it regularly?

- ○ yes
- ○ no
- ○ don't know

30c. How long, in total, have you taken this medication regularly? *If you started and stopped and then started again, please count only the time you were taking this medication.*

- ___ ___ number of months *or*
- ___ ___ number of years
- ○ don't know

Have you ever taken any of the following medications regularly (at least twice a week for more than a month)? *(continued)*

31. **Calcium pills or tablets**

- O yes
- O no → *Please go to #32*
- O don't know → *Please go to #32*

31a. How often did you usually take it when you were taking it regularly (that is, at least twice a week for more than a month)? *Please choose one of the following.*

- ___ ___ times per day *or*
- ___ ___ times per week
- O don't know

31b. **About two years** ago, were you taking it regularly?

- O yes
- O no
- O don't know

31c. How long, in total, have you taken this medication regularly? *If you started and stopped and then started again, please count only the time you were taking this medication.*

- ___ ___ number of months *or*
- ___ ___ number of years
- O don't know

32. **Calcium-based antacids (such as Tums, Rolaids, Extra-strength Rolaids, Alka-Mints, Chooz Antacid gum)**

- O yes
- O no → *If **female**, please go to #33 If **male**, please go to #44*
- O don't know → *If **female**, please go to #33 If **male**, please go to #44*

32a. How often did you usually take it when you were taking it regularly (that is, at least twice a week for more than a month)? *Please choose one of the following.*

- ___ ___ times per day *or*
- ___ ___ times per week
- O don't know

32b. **About two years** ago, were you taking it regularly?

- O yes
- O no
- O don't know

32c. How long, in total, have you taken this medication regularly? *If you started and stopped and then started again, please count only the time you were taking this medication.*

- ___ ___ number of months *or*
- ___ ___ number of years
- O don't know

Men: please go to #44 on page 17

Women: please continue with #33 on page 13

Menstruation, Pregnancy, and Menopause

33. How old were you when you had your **first** menstrual period?

__ __ years of age

O don't know

O never had a menstrual period

34. Have you ever been pregnant?

O yes

O no → *Please go to #35*

O don't know → *Please go to #35*

→ How many times have you been pregnant? *Please include miscarriages, stillbirths, tubal pregnancies and abortions.*

__ __ number of pregnancies

O don't know

34a. How many times were you pregnant with more than one baby (twins, triplets or more)? *If you are pregnant now, please do not include your current pregnancy.*

O never

__ __ number of pregnancies with more than one baby

O don't know

34b. How many of your pregnancies lasted 6 months or longer? (Pregnancy usually lasts 9 months. Six months is about the earliest a baby could survive.) *If you are pregnant now, please do not include your current pregnancy.*

O all of them

__ __ number of pregnancies lasting 6 months or longer

O don't know

34c. How many of your pregnancies resulted in live births?

O all of them

__ __ number of pregnancies with live-born children

O don't know

34d. How old were you at the **first** live birth?

age at **first** birth __ __ *or*

year of **first** birth __ __ __ __

O don't know

34e. How old were you at the **last** live birth?

age at **last** birth __ __ *or*

year of **last** birth __ __ __ __

O don't know

35. Have you ever used birth control pills or other hormonal contraceptives (implants or injections) for at least one year?

O yes

O no → *Please go to #36*

O don't know → *Please go to #36*

→ How old were you when you **first** used any of these hormonal contraceptives?

age at **first** use __ __ *or*

year of **first** use __ __ __ __

O don't know

35a. Were you still using hormonal contraceptives **about two years** ago?

O yes

O no

O don't know

35b. In **total,** how long did you take these hormonal contraceptives? *If you started and stopped and then started again, please count only the time you were taking these contraceptives.*

___ ___ number of years

O don't know

36. Have you had a **menstrual** period in the last 12 months? *Please include only* ***menstrual*** *bleeding, not bleeding that results from hormone replacement therapy (HRT) or progesterones, progestins or withdrawal bleeding.*

O yes → *Please go to #42*

O no

O don't know → *Please go to #42*

→ Have your periods stopped permanently or only temporarily due to pregnancy, breast-feeding, or other conditions?

O permanently

O temporarily → *Please go to #42*

37. How old were you when your periods stopped permanently?

age they stopped ___ ___ *or*

year they stopped ___ ___ ___ ___

O don't know

38. Why did your menstrual periods stop permanently? *Please tick all that apply.*

O natural menopause

O surgery

O radiation or chemotherapy

O other reason
Please specify: _____

O don't know

Please complete the next few questions which ask about surgeries you may have had.
Please answer all questions.

39. Hysterectomy (only the uterus or womb removed)

O yes

O no

O don't know

→ age when removed ___ ___ *or*

year when removed ___ ___ ___ ___

O don't know

39a. Hysterectomy with one ovary or part of an ovary removed

O yes

O no

O don't know

→ age when removed ___ ___ *or*

year when removed ___ ___ ___ ___

O don't know

39b. Hysterectomy with both ovaries removed

O yes

O no

O don't know

→ age when removed ___ ___ *or*

year when removed ___ ___ ___ ___

O don't know

39c. One ovary removed, completely or partly, without hysterectomy

O yes

O no

O don't know

→ age when removed ___ ___ *or*

year when removed ___ ___ ___ ___

O don't know

39d. Both ovaries removed without hysterectomy

- O yes
- O no
- O don't know
- → age when removed ___ ___ *or*
  year when removed ___ ___ ___ ___
- O don't know

40. **If you had radiation or chemotherapy**, when did you **first** have it?

- O had radiation or chemotherapy
- → age when this was given ___ ___ *or*
  year when this was given ___ ___ ___ ___
- O don't know
- O never had radiation or chemotherapy

41. **If your periods stopped permanently for any reason other than surgery, radiation or chemotherapy**, when did this occur?

- O other reason
  Please specify: _____
- → age when occurred ___ ___ *or*
  year when occurred ___ ___ ___ ___
- O don't know
- O not applicable

42. Doctors prescribe **hormone replacement therapy** for many reasons, including menopausal symptoms, surgical removal of the ovaries, osteoporosis, and heart disease prevention. (Menopausal symptoms include hot flashes, sweating, and depression.)

Have you ever taken hormone replacement therapy prescribed by a doctor and in the form of a pill or a patch?

Please do not include hormone therapy that was prescribed for birth control, infertility, hormone therapy delivered by injections, vagina creams or vaginal suppositories, or herbal or soy products.

- O yes
- O no → *Please go to #43*
- O don't know → *Please go to #43*

42a. Were you still having **menstrual** periods when you **first** took these hormones?

- O yes
- O no
- O don't know

42b. Were you prescribed either an estrogen-only pill or patch (such as Premarin) for hormone replacement therapy?

- O yes
- O no
- O don't know
- → How old were you when you **first** took estrogen-only medication?
  age when **first** taken ___ ___ *or*
  year when **first** taken ___ ___ ___ ___
  - O don't know

42c. Were you still using estrogen-only medication for hormone replacement therapy **about two years** ago?

- O yes
- O no
- O don't know

42d. In total, how long did you take estrogen-only medication for hormone replacement therapy? *If you started and stopped and then started again, please count only the time you were taking this medication.*

___ ___ number of months *or*

___ ___ number of years

- O don't know

42e. Progesterone or progestin is frequently prescribed by doctors together with estrogen for hormone replacement therapy. One common brand name is Provera. Another one is Prometrium. Have you ever taken progesterone or progestin together with estrogens for hormone replacement therapy?

- O yes
- O no → *Please go to #43*
- O don't know → *Please go to #43*

→ How old were you when you **first** took progesterone or progestin together with estrogens?

age when **first** taken ___ ___ *or*

year when **first** taken ___ ___ ___ ___

- O don't know

42f. Were you still using progesterone or progestin medication **about two years** ago?

- O yes
- O no
- O don't know

42g. In total, how long did you take progesterone or progestin together with estrogens? *If you started and stopped and then started again, please count only the time you were taking this medication.*

___ ___ number of months *or*

___ ___ number of years

- O don't know

43. Have you ever taken tamoxifen, raloxifene, or other anti-estrogen medication (such as Lupron or Depo-Provera)?

- O yes
- O no → *Please go to #44*
- O possibly - I have participated in a clinical trial for tamoxifen or other anti-estrogen medication
- O don't know → *Please go to #44*

→ What anti-estrogen medication did you take? *Please tick all that apply.*

- O tamoxifen
- O raloxifene
- O other: _____ *please specify*

43a. How old were you when you **first** took tamoxifen, raloxifene or other anti-estrogen medication?

age when **first** taken ___ ___ *or*

year when **first** taken ___ ___ ___ ___

- O don't know

43b. Were you still taking tamoxifen, raloxifene or other anti-estrogen medication **about two years** ago?

- O yes
- O no
- O don't know

43c. In total, how long did you take tamoxifen, raloxifene or other anti-estrogen medication? *If you started and stopped and then started again, please count only the time you were taking this medication.*

___ ___ number of months *or*

___ ___ number of years

- O don't know

Diet

44. **About two years ago,** on average, how often did you eat a piece or serving of **fruit**?

(A serving of fruit is: 1 medium-sized fresh fruit; 1/2 cup of chopped, cooked or canned fruit; 1/4 cup of dried fruit; 6 ounces of fruit juice (50%-100% pure juice).) *Please choose one of the following.*

__ __ servings per day *or*

__ __ servings per week *or*

__ __ servings per month

O don't know

45. **About two years** ago, on average, how often did you eat a serving of **vegetables**? *Please include green salads, beans, lentils, etc., and potatoes (not packaged potato chips).*

(A serving of vegetables is: 1 cup raw leafy vegetables;1/2 cup of other vegetables, cooked or chopped raw; 6 ounces of vegetable juice.) *Please choose one of the following.*

__ __ servings per day *or*

__ __ servings per week *or*

__ __ servings per month

O don't know

46. **About two years** ago, on average, how often did you eat a serving of **red meat** (not chicken or fish)?

(A serving of red meat is: 2-3 ounces of red meat (a piece of meat about the size of a deck of cards). Red meats include: beef, steak, hamburger, prime rib, ribs, beef hot dogs, beef-based processed meat, veal, pork, bacon, pork sausage, ham, lamb, venison.)
Please choose one of the following.

__ __ servings per day *or*

__ __ servings per week *or*

__ __ servings per month

O didn't eat red meat → *Please go to #47*

O don't know

46a. **About two years** ago, on average, how often did you eat a serving of **red meat** that was cooked by broiling, grilling, barbecueing or pan-frying (**not** stir-fried or deep-fried)? *Please choose one of the following.*

__ __ servings per day *or*

__ __ servings per week *or*

__ __ servings per month

O didn't eat red meat that was cooked by these methods → *Please go to #47*

O don't know

46b. On average, when you ate **red meat** cooked by these methods, which of the following best describes its appearance?

What was its **outside** appearance?	What was its **inside** appearance (how well done it was)?
○ lightly browned	○ red (rare)
○ medium browned	○ pink (medium)
○ heavily browned or blackened	○ brown (well-done)
○ don't know	○ don't know

47. **About two years** ago, on average, how often did you eat a serving of **chicken**? *Please do not include turkey or any other bird.*

(A serving of chicken is: 2-3 ounces of chicken meat; 1 drumstick; 1 thigh; half a breast; 2 wings; 3 nuggets.) *Please choose one of the following.*

___ ___ servings per day *or*

___ ___ servings per week *or*

___ ___ servings per month

○ didn't eat chicken → *Please go to #48*

○ don't know

47a. **About two years** ago, on average, how often did you eat a serving of **chicken** that was cooked by broiling, grilling, barbecueing or pan-frying (**not** stir-fried or deep-fried)? *Please choose one of the following.*

___ ___ servings per day *or*

___ ___ servings per week *or*

___ ___ servings per month

○ didn't eat chicken that was cooked by these methods → *Please go to #48*

○ don't know

47b. On average, when you ate **chicken** cooked by these methods, which of the following best describes its appearance?

What was its **outside** appearance?

○ lightly browned

○ medium browned

○ heavily browned or blackened

○ don't know

Physical Activity

We would like you to think back to when you were **in your 20s** and remember the physical activities you participated in then.

48. **In your 20s**, did you participate **regularly** in physical activity **for a total of at least 30 minutes a week**? *Please describe your activities below.*

		For how many years?	During those years, for how many months per year?	During those months, on average, for how many minutes or hours per week?
Walking	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Jogging (running slower than a mile in 10 minutes)	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Running (running faster than a mile in 10 minutes)	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Bicycling (including using an exercise bicycle)	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Swimming laps	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Tennis, squash racquetball	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Calisthenics, aerobics, vigorous dance (including ballet), **using a rowing machine, lifting weights**	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Football, soccer rugby, basketball	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Heavy household work (examples: using a non-power mower, shoveling, moving heavy loads, scrubbing floors)	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week

In your 20s, did you do any other **strenuous activities**? Strenuous activity means something that really increased your heart rate, made you hot, and caused you to sweat. Some examples are: skiing, skating, hockey, hunting, sledding or tobogganing, water-skiing.

Activity *please specify*	**For how many years?**	**During those years, for how many months per year?**	**During those months, on average, for how many minutes or hours per week?**
_____ →	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
_____ →	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
_____ →	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
_____ →	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
_____ →	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
_____ →	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week

49. When you were **in your 20s**, what was your usual occupation? (We mean what you did for the longest time, including any paid or unpaid employment, such as being a student or housewife or being unemployed.)

_____ occupation

○ don't know

If you are younger than age 31, please go to the next section (Alcohol Consumption) on page 25. Otherwise, please continue with #50.

Now, please think back to **your 30s and 40s.**

50. **In your 30 and 40s**, did you participate **regularly** in physical activity **for a total of at least 30 minutes a week**? *Please describe your activities below.*

		For how many years?	**During those years, for how many months per year?**	**During those months, on average, for how many minutes or hours per week?**
Walking	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Jogging (running slower than a mile in 10 minutes)	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Running (running faster than a mile in 10 minutes)	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Bicycling (including using an exercise bicycle)	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Swimming laps	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Tennis, squash racquetball	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Calisthenics, aerobics, vigorous dance (including ballet), **using a rowing machine, lifting weights**	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Football, soccer rugby, basketball	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Heavy household work (examples: using a non-power mower, shoveling, moving heavy loads, scrubbing floors)	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week

In your 30s and 40s, did you do any other **strenuous activities**? Strenuous activity means something that really increased your heart rate, made you hot, and caused you to sweat. Some examples are: skiing, skating, hockey, hunting, sledding or tobogganing, water-skiing.

Activity *please specify*	**For how many years?**	**During those years, for how many months per year?**	**During those months, on average, for how many minutes or hours per week?**
_____ →	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
_____ →	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
_____ →	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
_____ →	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
_____ →	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
_____ →	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week

51. When you were in your **30s and 40s**, what was your usual occupation? (We mean what you did for the longest time, including any paid or unpaid employment, such as being a student or housewife or being unemployed.)

_____ occupation

○ don't know

If you are younger than age 51, please go to the next section (Alcohol Consumption) on page 25. Otherwise, please continue with #52.

Now, please think back to **since you turned 50.**

52. **Since you turned 50**, did you participate **regularly** in physical activity **for a total of at least 30 minutes a week**? *Please describe your activities below.*

		For how many years?	During those years, for how many months per year?	During those months, on average, for how many minutes or hours per week?
Walking	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Jogging (running slower than a mile in 10 minutes)	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Running (running faster than a mile in 10 minutes)	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Bicycling (including using an exercise bicycle)	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Swimming laps	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Tennis, squash racquetball	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Calisthenics, aerobics, vigorous dance (including ballet), **using a rowing machine, lifting weights**	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Football, soccer rugby, basketball	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
Heavy household work (examples: using a non-power mower, shoveling, moving heavy loads, scrubbing floors)	○ yes → ○ no	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week

Since you turned 50, did you do any other **strenuous activities**? Strenuous activity means something that really increased your heart rate, made you hot, and caused you to sweat. Some examples are: skiing, skating, hockey, hunting, sledding or tobogganing, water-skiing.

Activity *please specify*	**For how many years?**	**During those years, for how many months per year?**	**During those months, on average, for how many minutes or hours per week?**
_____ →	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
_____ →	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
_____ →	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
_____ →	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
_____ →	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week
_____ →	___ ___ years	___ ___ months	___ ___ minutes per week *or* ___ ___ hours per week

53. **Since you turned 50**, what was your usual occupation? (We mean what you did for the longest time, including any paid or unpaid employment, such as being a student or housewife or being unemployed.)

_____ occupation

○ don't know

Alcohol Consumption

We would like you to think back to when you were **in your 20s.**

54. **In your 20s**, did you ever consume any alcoholic beverages at least **once a week for 6 months or longer**? *Please describe your consumption below.*

		For how many years?	**Dring those years, how much did you typically consume?**
Beer, hard cider (at least 3% alcohol)	○ yes → ○ no ○ don't know	___ ___ years consumed	___ ___ number of 12 ounce cans or bottles ○ per day ○ per week ○ don't know
Wine	○ yes → ○ no ○ don't know	___ ___ years consumed	___ ___ number of 4 ounce glasses of wine ○ per day ○ per week ○ don't know
Sake, sherry, port	○ yes → ○ no ○ don't know	___ ___ years consumed	___ ___ number of 1 ounce servings ○ per day ○ per week ○ don't know
Spirits, liquor mixed drinks, brandy, liqueurs	○ yes → ○ no ○ don't know	___ ___ years consumed	___ ___ number of 1 ounce shots liquor or spirits ○ per day ○ per week ○ don't know

55. When you were in your 20s, how many years **in total** did you consume **at least one alcoholic beverage (of any type) a week**?

___ ___ years consumed

○ never consumed alcohol

56. On average, how many alcoholic beverages a week did you consume during those years? That is, how many 4 ounce glasses of wine **or** 12 ounce cans or bottles of beer or hard cider, **or** 1 ounce servings of sake, sherry, port, or spirits, mixed drinks and cocktails.

___ ___ number of alcoholic beverages a week

○ never consumed alcohol

If you are younger than age 31, please go to the next section (Smoking) on page 28. Otherwise, please continue with #57.

Now, please think back to **your 30s and 40s.**

57. **In your 30s and 40s**, did you ever consume any alcoholic beverages at least **once a week for 6 months or longer**? *Please describe your consumption below.*

		For how many years?	During those years, how much did you typically consume?
Beer, hard cider (at least 3% alcohol)	O yes → O no O don't know	___ ___ years consumed	___ ___ number of 12 ounce cans or bottles O per day O per week O don't know
Wine	O yes → O no O don't know	___ ___ years consumed	___ ___ number of 4 ounce glasses of wine O per day O per week O don't know
Sake, sherry, port	O yes → O no O don't know	___ ___ years consumed	___ ___ number of 1 ounce servings O per day O per week O don't know
Spirits, liquor mixed drinks, brandy, liqueurs	O yes → O no O don't know	___ ___ years consumed	___ ___ number of 1 ounce shots liquor or spirits O per day O per week O don't know

58. When you were in your 30s and 40s, how many years **in total** did you consume **at least one alcoholic beverage (of any type) a week**?

___ ___ years consumed

O never consumed alcohol

59. On average, how many alcoholic beverages a week did you consume during those years? That is, how many 4 ounce glasses of wine **or** 12 ounce cans or bottles of beer or hard cider, **or** 1 ounce servings of sake, sherry, port, or spirits, mixed drinks and cocktails.

___ ___ number of alcoholic beverages a week

O never consumed alcohol

If you are younger than age 51, please go to the next section (Smokimg) on page 28. Otherwise, please continue with #60.

Now, please think back to **since you turned 50.**

60. **Since you turned 50**, did you ever consume any alcoholic beverages at least **once a week for 6 months or longer**? *Please describe your consumption below.*

		For how many years?	**During those years, how much did you typically consume?**
Beer, hard cider (at least 3% alcohol)	○ yes → ○ no ○ don't know	___ ___ years consumed	___ ___ number of 12 ounce cans or bottles ○ per day ○ per week ○ don't know
Wine	○ yes → ○ no ○ don't know	___ ___ years consumed	___ ___ number of 4 ounce glasses of wine ○ per day ○ per week ○ don't know
Sake, sherry, port	○ yes → ○ no ○ don't know	___ ___ years consumed	___ ___ number of 1 ounce servings ○ per day ○ per week ○ don't know
Spirits, liquor mixed drinks, brandy, liqueurs	○ yes → ○ no ○ don't know	___ ___ years consumed	___ ___ number of 1 ounce shots liquor or spirits ○ per day ○ per week ○ don't know

61. Since you turned 50, how many years **in total** did you consume **at least one alcoholic beverage (of any type) a week**?

___ ___ years consumed

○ never consumed alcohol

62. On average, how many alcoholic beverages a week did you consume during those years? That is, how many 4 ounce glasses of wine **or** 12 ounce cans or bottles of beer or hard cider, **or** 1 ounce servings of sake, sherry, port, or spirits, mixed drinks and cocktails.

___ ___ number of alcoholic beverages a week

○ never consumed alcohol

Smoking

63. Have you ever smoked at least one **cigarette** a day for 3 months or longer?
- O yes
- O no → *Please go to #64*
- O don't know → *Please go to #64*

63a. When did you **first** start smoking at least one cigarette a day?

age at **first** use ___ ___ *or*
year of **first** use
___ ___ ___ ___
- O don't know

63b. During periods when you smoked regularly, how many cigarettes did you typically smoke in a day?

___ ___ ___ cigarettes per day
- O don't know

63c. **About two years** ago, were you still smoking at least one cigarette a day?
- O yes
- O no
- O don't know

63d. Do you still smoke at least one cigarette a day?
- O yes → *Please go to #63f*
- O no → *Please go to #63e*
- O don't know

63e. When did you stop smoking at least one cigarette a day (we mean stop smoking permanently)?

age at last use ___ ___ *or*
year of last use
___ ___ ___ ___
- O don't know

63f. How many years, in total, did you smoke at least one cigarette a day for 3 months or longer? *(If you have stopped and restarted at least once, count only the time when you were smoking.)*

___ ___ total number of years
- O don't know

64. Have you ever smoked at least one **cigar** a month for at least 3 months?
- O yes
- O no → *Please go to #65*
- O don't know → *Please go to #65*

64a. When did you **first** start smoking at least one cigar a month?

age at **first** use ___ ___ *or*
year of **first** use
___ ___ ___ ___
- O don't know

64b. During periods when you smoked regularly, how many cigars did you typically smoke in a month?

___ ___ ___ cigars per month
- O don't know

64c. **About two years** ago, were you still smoking at least one cigar a month?
- O yes
- O no
- O don't know

64d. Do you still smoke at least one cigar a month?
- O yes → *Please go to #64f*
- O no → *Please go to #64e*
- O don't know

64e. When did you stop smoking at least one cigar a month (we mean stop smoking permanently)?

age at last use ___ ___ *or*
year of last use
___ ___ ___ ___
- O don't know

64f. How many years, in total, did you smoke at least one cigar a month for 3 months or longer? *(If you have stopped and restarted at least once, count only the time when you were smoking.)*

___ ___ total number of years
- O don't know

65. Have you ever smoked at least one **pipe** a month for at least 3 months?

- O yes
- O no → *Please go to #66*
- O don't know → *Please go to #66*

65a. When did you **first** start smoking at least one pipe a month?

age at **first** use ___ ___ *or*

year of **first** use

___ ___ ___ ___

- O don't know

65b. During periods when you smoked regularly, how many pipes did you typically smoke in a month?

___ ___ ___ pipes per month

- O don't know

65c. **About two years** ago, were you still smoking at least one pipe a month?

- O yes
- O no
- O don't know

65d. Do you still smoke at least one pipe a month?

- O yes → *Please go to #65f*
- O no → *Please go to #65e*
- O don't know

65e. When did you stop smoking at least one pipe a month (we mean stop smoking permanently)?

age at last use ___ ___ *or*

year of last use

___ ___ ___ ___

- O don't know

65f. How many years, in total, did you smoke at least one pipe a month for 3 months or longer? *(If you have stopped and restarted at least once, count only the time when you were smoking.)*

___ ___ total number of years

- O don't know

Height and Weight

66. About how tall are you, without your shoes on?

___ feet ___ ___ inches

or

___ ___ ___ centimetres

- O don't know

67. How much did you weigh **about two years ago**?

___ ___ pounds

or

___ ___ kilograms

- O don't know

68. How much did you weigh when you were **about 20 years old**?

___ ___ ___ pounds

or

___ ___ kilograms

- O don't know

Additional Information

69. Previous to this study, have you and your relatives ever taken part in any family health studies?

- O yes
- O no
- O don't know

Background Information

70. What is the highest level of education that you completed?

- ○ less than 8 years
- ○ 8 to 11 years
- ○ high school graduate
- ○ vocational or technical school
- ○ some college or university
- ○ bachelor's degree
- ○ graduate degree
- ○ don't know

71. Country of birth sometimes affects disease risk. Please fill in country of birth for **yourself, your parents and your grandparents**.

In addition, scientists have found that some genetic traits are more common or less common among Jewish people of different ethnic backgrounds. Please answer the questions about Jewish descent for each person.

	Country of birth	Is this person of Jewish descent?	Ashkenazi (East European)	Sephardic	other	don't know
You	_____	○ yes →	○	○	○	○
		○ no				
		○ don't know				
Your mother	_____	○ yes →	○	○	○	○
		○ no				
		○ don't know				
Your father	_____	○ yes →	○	○	○	○
		○ no				
		○ don't know				
Your mother's mother	_____	○ yes →	○	○	○	○
		○ no				
		○ don't know				
Your mother's father	_____	○ yes →	○	○	○	○
		○ no				
		○ don't know				
Your father's mother	_____	○ yes →	○	○	○	○
		○ no				
		○ don't know				
Your father's father	_____	○ yes →	○	○	○	○
		○ no				
		○ don't know				

72. How many years have **you** lived in Canada?

- ○ all my life
- __ __ number of years
- ○ don't know

73. Ethnicity and race sometimes affect disease risk. Scientists have found that some genetic traits are more common or less common among people of different backgrounds. We would like to know if this is true for genes associated with colorectal cancer.

Please fill in the background for **yourself, your parents and your grandparents**.
Please tick all that apply.

	You	Your mother	Your father	Your mother's mother	Your mother's father	Your father's mother	Your father's father
Black, from Africa	○	○	○	○	○	○	○
Black, from the Caribbean (e.g Trinidad, Jamaica, Haiti)	○	○	○	○	○	○	○
Black, from North America	○	○	○	○	○	○	○
Black, other	○	○	○	○	○	○	○
White	○	○	○	○	○	○	○
First Nations (e.g. Indian, Inuit)	○	○	○	○	○	○	○
North African (e.g. Egyptian)	○	○	○	○	○	○	○
Middle Eastern (e.g. Iranian)	○	○	○	○	○	○	○
Filipino	○	○	○	○	○	○	○
Japanese	○	○	○	○	○	○	○
Korean	○	○	○	○	○	○	○
Chinese	○	○	○	○	○	○	○
Other South East Asian (e.g.Vietnamese)	○	○	○	○	○	○	○
South Asian (e.g. East Indian, Pakistani)	○	○	○	○	○	○	○
other: *please specify*							
don't know	○	○	○	○	○	○	○

74. Which of the following categories best describes your total annual **household** income **about two years ago**?

- ○ no income
- ○ less than $6,000
- ○ $6,000 - $11,999
- ○ $12,000 - $19,999
- ○ $20,000 - $29,999
- ○ $30,000 - $39,999
- ○ $40,000 - $49,999
- ○ $50,000 - $59,999
- ○ $60,000 - $69,999
- ○ $70,000 - $79,999
- ○ $80,000 or more
- ○ don't know

75. In case we need to contact you in the future and you have moved, could we have the name of someone who is not living with you to whom we might write or call for your new address?

Name of relative or friend: _____

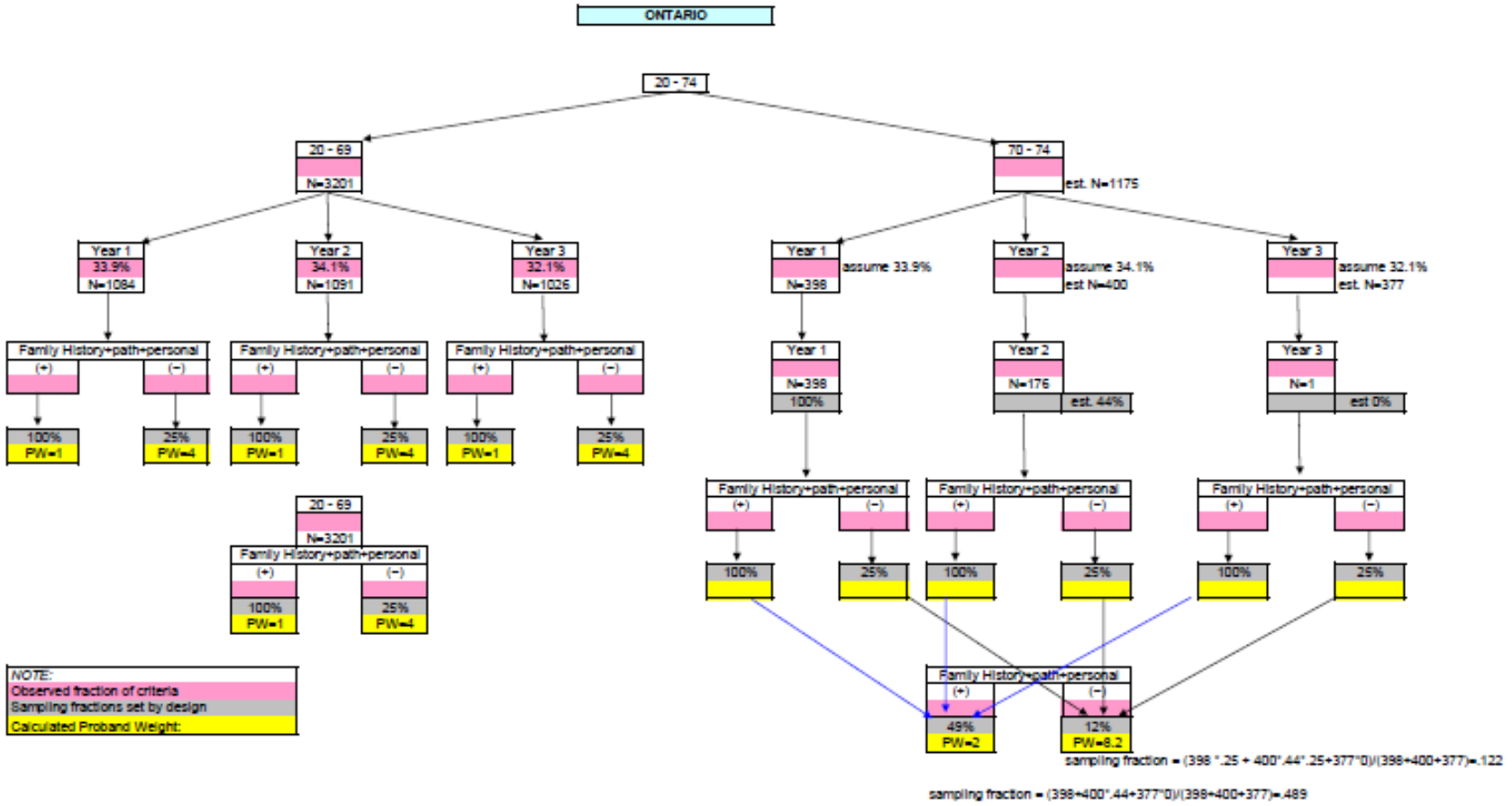
His or her address: _____

_____

His or her telephone number: (__ __ __) __ __ __ - __ __ __ __

Source: Cancer Family Registries – Informatics Support Centre.
http://www.coloncfr.org/images/questionnaires/Ontario/Ontario%20-%20Baseline.pdf

APPENDIX E – Sampling scheme and proband weight calculations from OFCCR



Sampling schematic of OFCCR received via email from D. Daftary, MD, M.H.Sc. Cancer Care Ontario. March 14, 2013.

APPENDIX F – Stratified sub-analysis

<40 group – Frequency tables

Sub-analysis – MALE (<40) – Comparison of Clinical sample by cases and controls			
Variable	**Division**	**Cases (N=59)**	**Controls (N=12)**
		N (%)	**N (%)**
BMI	Under/Normal	19 (32.20)	6 (50.00)
	Overweight	29 (49.15)	5 (41.67)
	Obese	11(32.20)	1 (8.33)
Physical Activity	0 hours/week	2 (3.39)	2 (16.67)
	>0 and ≤2 hours/week	7 (11.86)	1 (8.33)
	>2 and ≤4 hours/week	5 (8.47)	2 (16.67)
	>4 hours/week	45 (76.27)	7 (58.33)
NSAID use	Yes	1 (1.69)	0 (0.00)
	No	58 (98.31)	12 (100.00)
Vegetables	<5 Servings/day	58 (98.31)	12 (100.00)
	≥5 Servings/day	1 (1.69)	0 (0.00)
Smoking	Never Smoker	36 (61.02)	9 (75.00)
	>0 and <20 PYs	20 (33.90)	2 (16.67)
	20-29 PYs	2 (3.39)	1 (8.33)
	≥ 30 PYs	1 (1.69)	0 (0.00)
FH	0 FDRs with CRC	49 (83.05)	12 (100.00)
	1 FDRs with CRC	8 (13.56)	0 (0.00)
	2+ FDRs with CRC	34 (3.39)	0 (0.00)

BMI, body mass index; PYs, pack years; FH, family history; FDR, first-degree relative; CRC, colorectal cancer.

Sub-analysis – FEMALE (<40) – Comparison of Clinical sample by cases and controls			
Variable	**Division**	**Cases (N=67)**	**Controls (N=30)**
		N (%)	**N (%)**
BMI	Under/Normal	47 (70.15)	17 (56.67)
	Overweight	14 (20.90)	8 (26.67)
	Obese	6 (8.96)	5 (16.67)
Physical Activity	0 hours/week	6 (8.96)	4 (13.33)
	>0 and ≤2 hours/week	6 (8.96)	1 (3.33)
	>2 and ≤4 hours/week	6 (8.96)	5 (16.67)
	>4 hours/week	49 (73.13)	20 (66.67)
NSAID use	Yes	2 (2.99)	2 (6.67)
	No	65 (97.01)	28 (93.33)
Vegetables	<5 Servings/day	63 (94.03)	29 (96.67)
	≥5 Servings/day	4 (5.97)	1 (3.33)
Estrogen Status	Negative	0 (0.00)	0 (0.00)
	Positive	67 (100.00)	30 (100.00)
FH	0 FDRs with CRC	49 (73.13)	28 (86.11)
	1 FDRs with CRC	17 (25.37)	1 (3.33)
	2+ FDRs with CRC	1 (1.49)	1 (3.33)

BMI, body mass index; FH, family history; FDR, first-degree relative; CRC, colorectal cancer.

40 and over

Comparison of ≥40 and All-ages Multivariate-adjusted ORs

40 and over – MALE - Multivariate-adjusted ORs		
	≥40 OR (95% CI)	**All ages OR (95% CI)**
BMI		
Under/Normal	1.00	1.00
Overweight	1.892 (1.440, 2.486)	1.894 (1.450, 2.475)
Obese	2.238 (1.579, 3.171)	2.266 (1.608, 3.192)
Physical Activity		
0 hrs/wk	1.00	1.00
>0 and ≤2 hrs/wk	1.310 (0.795, 2.159)	1.363 (0.833, 2.229)
>2 and ≤4 hrs/wk	0.961 (0.567, 1.628)	0.980 (0.584, 1.644)
>4 hrs/wk	1.086 (0.714, 1.651)	1.135 (0.751, 1.713)
NSAID use		
No	1.00	1.00
Yes	0.546 (0.322, 0.925)	0.553 (0.328, 0.931)
Vegetables		
<5 Servings/day	1.00	1.00
≥5 Servings/day	1.490 (0.491, 4.520)	1.548 (0.519, 4.617)
Smoking		
Never Smoker	1.00	1.00
>0 and <20 PYs	1.185 (0.883, 1.590)	1.218 (0.914, 1.624)
20-29 PYs	1.380 (0.921, 2.067)	1.378 (0.925, 2.054)
≥ 30 PYs	1.206 (0.877, 1.659)	1.226 (0894, 1.681)
Age		
<40	-	1.00
40-44	1.00	0.577 (0.240, 1.388)
45-49	1.762 (0.899, 3.453)	1.015 (0.435, 2.368)
50-54	0.358 (0.197, 0.651)	0.207 (0.094, 0.456)
55-59	0.268 (0.151, 0.477)	0.154 (0.071, 0.335)
60-64	0.393 (0.224, 0.690)	0.226 (0.105, 0.485)
65-69	0.381 (00.218, 0.665)	0.219 (0.102, 0.468)
FH - FDRs with CRC		
0	1.00	1.00
1	2.37 (1.63, 3.46)	2.42 (1.66, 3.51)
≥2	5.35 (1.75, 16.39)	5.51 (1.81, 16.72)

OR, odds ratio; PYs, pack years; FH, family history; FDR, first-degree relative; CRC, colorectal cancer.

40 and over – FEMALE - Multivariate-adjusted ORs		
	≥40 OR (95% CI)	**All ages OR (95% CI)**
BMI		
Under/Normal	1.00	1.00
Overweight	1.160 (0.897, 1.501)	1.121 (0.873, 1.438)
Obese	1.102 (0.810, 1.499)	1.057 (0.783, 1.426)
Physical Activity		
0 hrs/wk	1.00	1.00
>0 and ≤2 hrs/wk	1.340 (0.842, 2.132)	1.384 (0.881, 2.175)
>2 and ≤4 hrs/wk	1.393 (0.891, 2.177)	1.359 (0.882, 2.094)
>4 hrs/wk	1.451 (0.991, 2.122)	1.457 (1.009, 2.104)
NSAID use		
No	1.00	1.00
Yes	0.479 (0.304, 0.755)	0.478 (0.307, 0.745)
Vegetables		
<5 Servings/day	1.00	1.00
≥5 Servings/day	0.777 (0.483, 1.250)	0.812 (0.511, 1.288)
Estrogen Status		
Negative	1.00	1.00
Positive	1.030 (0.801, 1.324)	0.969 (0.755, 1.244)
Age		
<40	-	1.00
40-44	1.00	0.852 (0.461, 1.576)
45-49	2.124 (1.253,3.600)	1.815 (1.001, 3.293)
50-54	0.787 (0.481, 1.288)	0.672 (0.382, 1.184)
55-59	0.724 (0.445, 1.178)	0.619 (0.354, 1.085)
60-64	0.763 (0.481, 1.209)	0.655 (0.383, 1.121)
65-69	0.920 (0.576, 1.479)	0.788 (0.458, 1.357)
FH - FDRs with CRC		
0	1.00	1.00
1	2.37 (1.63, 3.46)	2.42 (1.66, 3.51)
≥2	5.35 (1.75, 16.39)	5.51 (1.81, 16.72)

OR, odds ratio; PYs, pack years; FH, family history; FDR, first-degree relative; CRC, colorectal cancer

Comparison of AUCs for the ≥40 and All-ages sample

Comparison of ≥40 and all ages AUCs		
Model	**≥40 AUC**	**All ages AUC**
Male		
Base	0.7094	0.7246
FH - Included	0.7502	0.7623
Female		
Base	0.6157	0.6214
FH - Included	0.6219	0.6275