How effective are colorectal cancer screening programs at increasing the rate of screening in asymptomatic average-risk groups in Canada?

Thomas James Charters
Department of Epidemiology, Biostatistics and Occupational Health
McGill University, Montréal
August 2011
A thesis submitted to McGill University in partial fulfillment of the degree of Master of Science in Epidemiology

Table of Contents

Abstract

Colorectal Cancer (CRC) is the third most commonly diagnosed cancer and second highest cause of cancer related mortality in Canada. Despite availability of screening services and establishment of guidelines, utilization of screening procedures in Canada has historically been low. *ColonCancerCheck* is an organized colorectal cancer screening program introduced in Ontario in 2008 which aims to increase screening adherence. The objectives of this study are to estimate the impact of *ColonCancerCheck* on screening behavior in the overall asymptomatic average risk population, examine demographic predictors of screening, and investigate how these are modified. Additionally, this analysis measures the effectiveness of *ColonCancerCheck* in how it modifies age-specific screening rates at the recommended initiation age of 50 years.

This analysis uses data from five cycles of the Canadian Community Health Survey prior to and following program implementation. Survey cycles were pooled to create one average pseudo-population and bootstrap repeated replication techniques were applied for accurate variance calculation. A difference-in-differences design was used to evaluate the overall impact of *ColonCancerCheck* in Ontario relative to the rest of Canada and a regression discontinuity design was used to measure changes in screening rates at age 50. Outcomes include self-report of guaiac fecal occult blood test in the previous year or colonoscopy or flexible sigmoidoscopy in the previous year.

Results indicate that factors which consistently influence the likelihood of screening include being physically active, having taken a flu shot, being of an older age, and having a regular medical doctor and greater numbers of physician consultations. The difference in differences analysis indicated that *ColonCancerCheck* has significantly increased screening in the average risk population although there is insufficient evidence to say that it has altered the demographic predictors of screening. Additionally, results from the regression discontinuity design indicate that screening rates at the age threshold of 50 in Ontario following the implementation of *ColonCancerCheck* increased significantly although further analysis was inconclusive as to the program causing this increase.

Abrégé

Le cancer colorectal est le 3e cancer le plus communément diagnostiqué et le 2e type de cancer le plus mortel au Canada. Malgré l'existence de services de dépistage et l'établissement de directives, l'utilisation des procédures de dépistage a toujours été faible au Canada. *ContrôleCancerColorectal* est un programme structuré de dépistage de cancer colorectal mis en place en Ontario en 2008 et dont la finalité est d'augmenter la participation au dépistage. Cette étude a pour objectifs d'évaluer les répercussions de *ContrôleCancerColorectal* sur le comportement face au dépistage de l'ensemble de la population asymptomatique à risque moyen, d'examiner les prédicteurs démographiques de dépistage et d'analyser comment ceux-ci sont modifiés. Cette étude mesure également l'efficacité de *ContrôleCancerColorectal,* à savoir comment celui-ci modifie les taux de dépistage par âge à l'âge initial recommandé de 50 ans.

Cette analyse utilise les données de cinq cycles de l'Enquête sur la santé dans les communautés canadiennes, avant et après la mise en place du programme. Les cycles d'enquête ont été regroupés afin de créer une pseudo-population moyenne et des méthodes de ré-échantillonnage bootstrap ont été adoptées pour effectuer un calcul précis de la variance. Un modèle de différence dans les différences a été utilisé pour évaluer les répercussions générales de *ContrôleCancerColorectal* en Ontario par rapport au reste du Canada et un plan de discontinuité de la régression a permis de mesurer les changements des taux de dépistage à 50 ans. Les résultats incluent les rapports des patients sur les tests guaiac de recherche de sang occulte dans les selles au cours de l'année passée, ou une colonoscopie ou une sigmoïdoscopie flexible au cours de l'année passée.

Les résultats indiquent que les facteurs qui influent de façon constante sur la probabilité de dépistage sont notamment le fait d'être physiquement actif, d'avoir reçu un vaccin contre la grippe, d'être plus âgé, d'avoir un médecin habituel et d'aller à un plus grand nombre de consultations. L'analyse de la différence dans les différences indique que *ContrôleCancerColorectal* a considérablement augmenté le taux de dépistage dans la population à risque moyen bien qu'il n'existe pas de preuves suffisantes démontrant qu'il ait changé les prédicteurs démographiques de dépistage. De plus, les résultats du plan de

discontinuité de la régression indiquent qu'en Ontario, à la suite de la mise en place de *ContrôleCancerColorectal,* les taux de dépistage à l'âge limite de 50 ans ont augmenté de façon significative bien qu'une analyse additionnelle n'ait pas permis de conclure quel était le programme à l'origine de cette augmentation.

Acknowledgements

I would like to use this section to demonstrate thanks and gratitude for the support of several individuals for whom I am indebted to their advice, support, and patience. Firstly, I would like to thank my supervisor Erin Strumpf for motivating the idea of this project and expanding the basis of my knowledge in epidemiology to realize the possibilities of new approaches to examining the impact of health services. Further thanks are required for offering the advice and encouragement on the research, writing, and statistical analysis relevant to this project to allow me to make the most out of this experience. Recognition is also required of my committee member Maida Sewitch who has offered substantive positive contributions in terms of her vast knowledge on colorectal cancer screening. I should also give thanks to members of the Epidemiology faculty including Sam Harper and Jay Kaufman for taking the time out of their busy schedules to answer methodological and technical questions which have increased by confidence in the validity and interpretability of the results I have presented here. Linda Rabeneck also contributed through particular knowledge of the timeline for the development of the screening program which was very instrumental in this project.

I also would like to voice my appreciation to the laboratory technicians Danielle Forest and Marie Eve-Gagnon at the Quebec Inter-University Centre for Social Statistics who have offered me technical advice and have always worked hard and quickly to approve my results for release in time. My thanks also goes to Andre Yves-Gagnon and Suzanne Lariviere in the Epidemiology administrative offices for their tireless support. Finally, I thank my parents for helping give me the opportunity to study at this institution.

I have been supported financially during this time by a Frederick Banting and Charles Best Canadian Institute of Health Research Canada Graduate Scholarship ($17 500) and by a Quebec Inter-university Centre for Social Statistics Matching Grant Award ($4 000).

Introduction

Among the challenges facing the Canadian healthcare system are those which concern the incidence and mortality derived from colorectal cancer. Cancers of the colon and rectum, or colorectal cancer (CRC), contribute to the second highest cause of cancer deaths in Canada and are the third most commonly diagnosed in men and women (1). The impact of this form of cancer on the healthcare system is augmented by the fact that within the previous three decades there has been a 117% increase in new cases and a 55% increase in colorectal cancer mortality largely thought to relate to a shift in population demographics to older ages (2). Furthermore, an observed tendency to diagnose this cancer at advanced stages in Canada (3) has contributed to lower probabilities of survival in those diagnosed (4) and greater costs to the healthcare system (5).

A potential means to address these problems has been the adoption of colorectal cancer screening guidelines based on existing knowledge of the importance of early detection of cancers (6-8). CRC screening exploits the window of opportunity in which cancers or pre-cancerous polyps may be detected and removed at the asymptomatic or latent stage (9) which greatly improves survival from colorectal cancer (4).

Several tools for colorectal cancer screening have been adopted to improve early detection of cancerous lesions. The guaiac fecal occult blood test (gFOBT) is one form of stool-based screening test that detects increases in blood in stools which accompany large adenomatous polyps or cancers (10, 11). Although this test is advantageous due to being non-invasive and requiring minimal preparation, it is insufficiently accurate to be used on its own (12). As such, positive results from this test are followed-up with endoscopic tests such as colonoscopy and flexible sigmoidoscopy which are able to make structural examinations of the entire bowel and simultaneously remove polyps if necessary (13). Several studies have proven the effectiveness of endoscopic procedures in reducing colorectal cancer mortality (14-17). Furthermore, several large population-based randomized trials have established that mass screening strategies employing guaiac fecal

occult blood tests with endoscopy follow-up on age-eligible segments of populations have significant effects on reducing overall colorectal cancer mortality (18).

In spite of these encouraging findings, screening for colorectal cancer in Canada has historically been low. In 2003, screening adherence for gFOBTs in average risk individuals was found to only be 15% (19). Evidence suggests that both patients (20) and physicians (20, 21) were unknowledgeable about proper screening protocols and tools.

In order to address problems relating to low screening compliance, the Ministry of Health and Long Term Care in Ontario launched an organized population-based colorectal cancer screening program called *ColonCancerCheck* in March of 2008 (22). This program followed previously made screening recommendations with gFOBT testing biennially and follow-up endoscopic tests. Furthermore, it involved provisions for testing in those without regular doctors, formation of registries for invitations and result letters, laboratory contracts for processing tests and meeting quality standards, and ongoing educational and media campaigns for physicians and patients (22). Although some sources have established increases in screening services in Ontario after the introduction of *ColonCancerCheck* (22, 23), this is insufficient evidence to make a claim as to the effect of the program since proportions screened were also observed to increase in several other provinces without screening programs at this time (23).

The purpose of this thesis is to measure and qualify the impact of *ColonCancerCheck* on colorectal cancer screening services in the asymptomatic average risk population. I first thoroughly review the literature in regards to rationale and justification for screening for colorectal cancer in Canada, discuss the evidence for the effectiveness of screening tools, and elaborate further on the development and strategies of *ColonCancerCheck.* Using the Canadian Community Health Survey, I examine demographic predictors of screening and utilize several quasi-experimental designs to estimate the effect of *ColonCancerCheck* in a causal framework. I analyze the absolute increases in screening in a Difference in Differences model, the change in characteristics which affect screening upon introduction of the program in a Difference in Difference in Differences model, and use a Regression Discontinuity Design to address whether the program impacts the screening behavior of individuals as they enter the average risk population at age 50.

Section I: Rationale for Screening

The general aim of early disease detection is to reduce mortality and morbidity due to a disease which can be intervened on in its early stage prior to clinical manifestation (9). The rationale for justifying screening derived from the classic World Health Organization (WHO) paper by Wilson and Jungner is as follows (9):

1. The condition sought should be an important health problem
2. There should be an accepted treatment for patients with recognized disease
3. Facilities for diagnosis and treatment should be available
4. There should be a recognizable latent or early symptomatic stage
5. There should be a suitable test or examination
6. The test should be acceptable to the population
7. The natural history of the condition, including development from latent to declared disease should be adequately understood
8. There should be an agreed policy on whom to treat as patients
9. The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole
10. Case finding should be a continuing process and not a "once and for all" project

These criteria have proven remarkably reliable over time in regard to a number of health challenges. Some effort has been made to update these criteria, mostly derived from a refined focus in medical ethics, program evaluation and scientific evidence to support the effectiveness and safety of tests and procedures (24), most of which is discussed later.

Condition 1

This condition states that the disease in question should be an important health problem. Combined, cancers of the colon and rectum have the third highest incidence of cancers and contribute to the second highest number of deaths derived from cancer in Canada in males and females combined. In 2010, the age standardized incidence rate of colorectal cancer (CRC) was estimated to be 62 per 100,000 in males and 41 per 100,000 in females.

Estimates state that 22 500 Canadians were diagnosed with CRC and that 9100 Canadians died from CRC in the year 2010 (1). Likely as a result of earlier detection and improved treatment, age-standardized CRC incidence rates have remained stable since the mid 1990's and CRC mortality rates have been gradually declining since before 1977, although the absolute health burden of CRC has increased largely due to rises in the size of the population and older age demographics (2). With Ontario as an example, the unadjusted incidence and mortality rates of colorectal cancer have risen over the previous three decades indicating a total increase in new cases of 117% and mortality by 55% (2). Although CRC is a prominent health concern throughout Canada, CRC incidence rates vary considerably by province. In 2005, British Columbia had the lowest colorectal cancer age standardized incidence rate at 44.9 per 100 000 whereas Newfoundland and Labrador had a rate of 67.3 per 100 000 with the Canadian average being 50.2 per 100 000 and Ontario having a similar incidence rate of 49.3 per 100 000 (25).

A related concern is the stage of cancer at which CRC is first diagnosed. Average lifetime costs of disease management for colon cancer increased from $20 319 in stage I cancer to $35 841 for stage IV cancer with results ranging from $27 505 for stage I rectal cancer to $36 939 for stage IV rectal cancer in 1998 $CAD (5). Most worrisome is the estimated probability of survival for late-stage detection of cancers. Commonly cited survival estimates from the Surveillance Epidemiology, and End Results Program (SEER) in the US indicate that five-year relative survival for individuals with, at first diagnosis, stage I cancer is 96%, stage II 87%, stage III 55%, and stage IV 5%, with only minor variation by location of cancer (4). These stages are related to the extent of cancer growth and invasion into surrounding tissues as based on the guidelines of the American Joint Committee on Cancer where higher numbered stages indicate increased cancer invasion (26). In the province of Manitoba in 2007, prior to the launch of any organized colorectal cancer screening program, only 20% of cancers were initially diagnosed in stage I and 32% in stage II. 24% of cancers were found to be initially diagnosed in stage III and 19% in stage IV, with 4% unknown (3). This is indicative of a problem in which CRC poses a substantial risk to population health particularly through late stage detection of invasive cancers.

Condition 2

This condition instructs that there should be an accepted treatment for individuals with the disease. Although this is not the focus of this report, I will very briefly describe the basic treatment strategy. Removal of cancerous lesions through surgery is noted to be the definitive component of treatment. Hemicolectomy, removal of the cancer with a section of the colon, is typically used for colon cancers whereas mesorectal excision is used in treatment of rectal cancers. Adjuvant chemotherapy is typically recommended for colon cancer surgery follow up with adjuvant radiotherapy being an additional option for those with rectal cancer (27). There is also the possibility of removing pre-cancerous polyps from the colon (see *Condition 4*). Polypectomy is the process of surgically removing polyps from the colon which can be accomplished with the assistance of colonoscopy. Polypectomy of larger polyps can be accomplished with a cautery snare and removal of smaller polyps with biopsy forceps. Histological examination of removed polyps and follow up of patients, given that subsequent polyp development after surgery has a likelihood of 30%, are other noteworthy aspects of the polypectomy procedure (28).

Condition 3

This condition states that facilities for diagnosis and treatment of the disease should be available. Prior to the launch of *ColonCancerCheck* in Ontario, services across Canada integral to the diagnosis and treatment for CRC were offered albeit on a non-systematic and individualistic basis (23). To address the anticipated increase in demand for tests and procedures associated with *ColonCancerCheck,* the government of Ontario also made additional funding available for the program (22).

Condition 4

One of the most important conditions in these criteria is the necessity for a latent or early symptomatic stage of the disease. This is a stage in which disease is present and progressing, but presents no symptoms (9). The adenoma-carcinoma sequence has been examined closely as far back as the 1960s when it became seriously hypothesized that benign polyps lead to intestinal cancers (29). More recently, it has been estimated that as many as 95% of colorectal cancers in Western populations were originally derived from adenomatous polyps (30). Data from the National Polyp Study has estimated an

approximate ten year timespan for a small adenomatous polyp to develop into symptomatic cancer, which provides a large window of opportunity for screening in the latent stage (31). It is worth clarifying that not all polyps indicate this level of concern as hyperplastic polyps are largely harmless and unlikely to develop into carcinomas (32). Histological analyses have found further differences in that small right-sided polyps in younger (<60 years) patients have an estimated 3.8% chance of containing advanced pathological features whereas polyps which are large (>1.0cm), left-sided, in older patients, that occur in the presence of anemia, or either as single or combined parameters of these, have a maximum predictive value of 75.4% for advanced adenomas (33).

Condition 5

This condition specifies that there should be a suitable test or examination. As will be discussed in later sections, several organizations in Canada (6-8) have made recommendations for screening for colorectal cancer with proven screening tools. Briefly, guaiac fecal occult blood tests (gFOBT) are widely agreed upon as the initial test for mass-screening of the average risk population. More invasive endoscopy tests such as the colonoscopy and the flexible sigmoidoscopy (FS) procedures are used to more precisely screen in advanced risk groups, including those who received a positive gFOBT result. Section II will further elaborate on the efficacy, use, and risks associated with these procedures.

Condition 6

This condition specifies that the screening test should be acceptable to the population. Generally, FOBT style tests are seen to be beneficial in that they are non-invasive, can be completed at patients' homes, and can be completed without the involvement of healthcare professionals (34). Early results from 1984 indicate that an Australian general population aged 40-74 had a compliance rate of 67% for gFOBT and that compliance was improved with written reminders although decreased in older individuals (35). Further evidence for gFOBT compliance is provided by a series of population-based randomized trials in which large segments of the general population were mailed screening kits. Briefly, initial compliance, which does not represent compliance for all screening invitations, ranged from approximately 60% in the Nottingham UK study (36) to 63% in

the Goteburg Sweden trial (37) to 67% in the Funen Denmark trial (38). An Australian survey on beliefs pertaining to gFOBT found that, despite awareness of the test, many individuals still felt no intention to screen. Predictors of screening included having a family history of CRC, beliefs that bowel cancer is curable when detected at an early stage, a perception of susceptibility to CRC, and an acceptance of the technique (39).

More recent information on screening indicates elderly individuals are more likely to adhere to screening guidelines which may relate to increased public awareness of CRC and commonality of screening (23). Evidence also suggests that greater involvement from the physician, even just through written support, is successful at increasing screening rates when compared to standard screening invitations (40). Colonoscopy, based on questionnaires for preferred test, has been shown to be the most acceptable of invasive screening tests due in large part to its accuracy (41). Colonoscopy with sedation was approximately twice as likely to be acceptable to patients compared to unsedated FS based on numbers of patients indicating they would return for rescreening (42). Other measures of acceptability including Visual Analogue Scales, pain, tolerance, satisfaction, and embarrassment have been measured for colonoscopy and found to be very high in many patients who commented that the procedure was not as unpleasant as anticipated (43). It is worth adding the caveat though that choice of screening test does not appear to significantly improve screening adherence (43).

Condition 7-8

Condition 7 maintains that the natural history of the disease, including development from latent to symptomatic disease should be adequately understood and *Condition 8* states that there should be an established policy on whom to treat as patients. I have outlined the major points of interest in the adenoma-carcinoma sequence in *Condition 4* which indicates that there is an in-depth understanding of the natural history of colorectal cancer. There also seems to be some agreement as to who constitutes the average risk patient population (6-8). One of the main points of contention is whether the screening program should target a patient population 50 years of age and older, as is the recommendation with the Canadian Task Force on Preventive Healthcare (7) and the Canadian Association of Gastroenterology (8), or a population of individuals aged 50-74 as is the

recommendation of the National Committee on Colorectal Cancer Screening (6). It is not absolutely clear why the latter organization capped the age range for screening, although 50-74 was the principal age range of the Population Health Model (POHEM) based on previous studies of screening effectiveness which the Committee derived its recommendations. Additionally, simulations found that screening for CRC conferred no significant increases to life expectancy if screening was performed beyond the age of 80 (44). For further reference on the natural history of colorectal cancer see (27, 31).

Condition 9

This condition states that the combined screening and treatment regimen should be cost effective. Reviews on this topic consistently support this criterion (12, 45, 46). One systematic review (46) of American studies found that screening of average risk individuals consistently cost less than $US 50 000 per life-year gained in comparison to no screening. The cost effectiveness of annual gFOBT was estimated in a range from $US 5691 to $US 17 805 per life-year gained and a regime of annual gFOBT with FS every five years estimated at $US 13792 to $US 22 518 per life-year gained. When comparing alternate treatment strategies there was no consensus as to the approach with the optimal incremental cost effectiveness ratio, although colonoscopy at 10 year intervals or gFOBT with FS were both highlighted in several studies as the most effective screening strategies. This systematic review included no examples of gFOBT with follow up colonoscopy. The authors believe that the cost effectiveness results compare favourably with other established screening programs such as mammography of women >50 years of age and treatment of moderate hypertension (46). Cost effectiveness estimates from the UK Nottingham based RCT (36) indicate that the additional costs of participation in screening compared to symptomatic presentation were £5290 per cancer detected and, under conservative assumptions, £1584 CI [£717, £8612] per life-year gained (47). Early estimates in the Canadian context using a microsimulation model have anticipated that biennial screening on 67% of individuals aged 50-74 in the year 2000 would result in a measure of $CAD 11 907 per life-year gained (48). Notably, due to screening costs, population preferences, and resource capacity, cost effectiveness results may differ between countries with European and Asian studies typically finding more favourable cost effectiveness results for CRC screening. However, for the Canadian system it has

been argued that estimates from the US and Australia are largely similar and are valid approximations for cost effectiveness in Canada (45).

Condition 10

This condition claims that case finding must be a continuous process rather than singular occurrence. This concern relates to the subsequent occurrence of polyps even after polypectomy (28) and is very much in agreement with all of the major recommendations for screening strategies offered in Canada and elsewhere. The Canadian Task Force on Preventive Health Care has recommended annual or biennial screening with gFOBT with follow up of positive results (7) as has the National Committee on Colorectal Cancer Screening (6) and the Canadian Association of Gastroenterology has recommended biennial screening of gFOBT, five year screening with FS or 10 year screening with colonoscopy (8).

Section II: Screening Tests

A. Stool-Based Screening Tests

Methods of screening for colorectal cancer (CRC) which utilize at-home stool tests are commonly used and often seen as advantageous due to their non-invasive nature, the convenience of being performed at home, and not requiring bowel preparation (13, 34, 49). The principal varieties of stool testing include the guaiac fecal occult blood test (gFOBT), the fecal immunochemical test (FIT) and the stool DNA (sDNA) test (34). Here I will be discussing the guaiac fecal occult blood test (gFOBT), which is the principal variety of the fecal occult blood test relevant to this project.

Occult gastrointestinal bleeding relates to bleeding that is not openly evident to an individual (10). Blood loss from the gastrointestinal tract typically ranges from 0.5ml to 1.5ml per day which is usually not detected by occult blood testing and blood loss of 100ml a day may still not be enough to alter the appearance of stools (10). Many fecal occult blood tests are used to test for presence of cancer under the assumption that large adenomatous polyps and early stage (Duke's A or B) cancers will bleed at detectable

levels (10, 11). However, in many cases asymptomatic cancers will vary considerably in the amount of fecal blood they release and may even indicate normal levels of fecal blood (49). Other factors which may influence the level of fecal occult blood include the level of bleeding, hemoglobin degradation by flora in the colon, stool transit time and storage, and stool mixing (49). Guaiac-based fecal occult blood tests make use of the fact that guaiac turns blue after oxidation by oxidants or peroxidases after implementation of a developing agent oxygen donor such as hydrogen peroxide (10). The heme in fecal blood acts as a pseudo-peroxidase which releases oxygen from hydrogen peroxide and reacts with the colourless guaiac paper to form a blue dye (10, 11). Common procedures for fecal occult blood testing include taking two distinct stool samples from each of three separate bowel movements onto three test cards with windows lined with guaiac paper, where a positive result occurs generally if there is a reaction in one or more of the windows (11, 12). Due to the concern that colorectal neoplasms may bleed sporadically and that the blood may not be spread consistently throughout a given bowel movement this creates the need to take multiple samples from different locations (11).

The likelihood of a fecal test indicating a positive result is often related to the quantity of fecal heme and the location of the bleeding (10). For instance, when using a fecal occult blood test (Haemoccult test) to examine fecal specimens from patients with colonic polyps with equivalent levels of bleeding (2.0-3.99 ml daily), the rate of positive tests were much higher in patients with polyps in the descending colon and rectosigmoid (86%) than in patients with polyps in the ascending and transverse colons (26%) (50). There has been concern that false positive results from gFOBT may result from a diet which includes foods with peroxidase activity (heme in red meat and certain fresh fruits and vegetables) or certain drugs such as non-steroidal anti-inflammatory drugs which can increase gastrointestinal bleeding (51). Concerns over false negatives derived from high levels of Vitamin C consumption are also common (51). If stool smears are too thick, or if they dry out or are exposed to high ambient temperatures, false negatives may also result (51). A meta-analysis of five randomized trials testing the effects of dietary constraints on screening found that such restrictions, if severe, may decrease test completion and that positivity rates are not affected by dietary restrictions (52) although using positivity as a

surrogate for test accuracy has been noted as being a limitation (52) in addition failing to address differences in diet across cultural groups (12).

Rabeneck et al. (53) take note that both the NHS Bowel Screening Program in the UK and the National Israeli Breast and Colorectal Cancer Detection Program do not require dietary restrictions, although in the UK diet restrictions may be requested for a second test if it is suspected that dietary factors lead to a false positive during the first test. The Finnish Cancer Registry is stricter on diet as participants are asked to restrict consumption of raw meat, blood, and liver three days prior to and during testing (53). gFOBT kits available in Canada typically recommend restrictions in diet up to three days before sample collection for red meats and liver, and in some cases restrictions for raw fruits and vegetables which contain peroxidase-like substances (53). There are inconclusive results from the literature on medication use on effecting gFOBT positivity (53). Some prospective studies have indicated that regular aspirin or NSAID use may have no effect on true positive results (54, 55) although another has indicated false positives attributable to medication use (56). Further uncertainty has risen due to methodological limitations in several of these studies related to the recruitment of healthy volunteers (53). As such, the panel for Cancer Care Ontario made no recommendations for prohibiting aspirin or NSAID use, although recommended restriction of Vitamin C supplements (53).

Sensitivity and specificity of the gFOBT can vary according to the test type and other conditions (12, 34, 53). Rehydration of samples as a means to improve test performance has generated some controversy due to its capability of increasing sensitivity in some tests, but a tendency to decrease specificity, decrease positive predictive values, and hurt test readability simultaneously (34). It has been found that the sensitivity of gFOBT increases significantly when three consecutive stool samples are used (57). Additionally, the positive predictive value of the test has been demonstrated to increase in regards to detection of colorectal cancer and advanced adenomatous polyps given greater numbers of positive slides (58, 59).

A large systematic review of guaiac fecal occult blood tests in average risk populations was undertaken by the Centre of Reviews and Dissemination which included 59 studies

containing information on diagnostic accuracy (12). This review found that no single guaiac fecal occult blood test was clearly superior out of the five studied which included Haemoccult, Haemoccult II (which were treated as a single test), KryptoHaem, Hemoccult Sensa, and Shionogi B. Reference standards differed between studies and included colonoscopy, barium enema, FS, follow up through a cancer registry, and combinations of these (12). In detection of all neoplasms in cohort studies, results for sensitivity ranged widely from 6.2% to 85% dependent on test preparation parameters such as whether the gFOBT was hydrated. Results from diagnostic case control studies typically reported higher sensitivities ranging from 47%-72% for detection of neoplasms in general and 25%-96% for detection of colorectal cancer. Specificities typically showed less variation and ranged from 87%-99% based on the test and the target. The analysts found that, in the case of diagnostic cohort study designs, Krytohaem had the highest sensitivity for detection of all neoplasms and non-rehydrated Haemoccult and Haemoccult Sensa had the highest recorded sensitivities for detecting CRC.

Very high and significant (Cochrane $Q<0.5$ and or $I^2>75\%$) heterogeneity measures of the results were found indicating both statistical and clinical differences. These resulted from disparities in study design, quality measures, test preparatory measures, patient populations, and use of different reference standards. The authors concluded that pooling of results for cumulative estimates was inappropriate (12). Another systematic review has found similarly disparate results for test accuracy between studies (53). In several cases, suspicion was drawn from the highest performing test results given the methodological limitations of case control studies and other problems in regard to reporting selection criteria and follow-up (60) and use of FS as the reference standard (61)[1].

It may be useful, instead of focusing on the whole body of literature, to identify the most high quality studies in order to obtain a better picture of the accuracy of gFOBT tests. In their systematic review, The Centre for Reviews and Dissemination (12) identifies what they consider to be the two highest quality diagnostic cohort studies based on

[1] As flexible sigmoidoscopy can only examine the distal section of the bowel, it is not adequate to detecting adenomas in the proximal bowel which results in upwards biased estimates of sensitivity when hard to detect adenomas are discounted. See (12, 65).

appropriately defining the patient population, and using colonoscopy as the reference standard in cases of having both positive and negative results. One study recruited 505 average risk participants in China and found that the unrehydrated Haemoccult test had a sensitivity of 19.1% 95% CI [13.4, 26.4] and specificity of 79.6% [74.9, 83.6] in detecting all colonic neoplasms and a sensitivity of 30.0% [0.6, 80.6] and specificity of 80.0% [76.3, 83.5] for detection of CRC (62). The second study went through 13 Veterans Affairs medical centers in the United States and recruited 3121 asymptomatic patients for rehydrated Haemoccult gFOBT (63). The authors found that sensitivities/specificities for detection of advanced neoplasia were 23.9%[19.0, 28.9]/93.9% [92.9, 94.8], for all neoplasms 11.3% [9.4, 13.5]/93.9%[92.6, 95.0], and for CRC 34.1%[22.8, 46.3]/92.6%[91.5,93.6] (63).

B. Flexible Sigmoidoscopy

Flexible Sigmoidoscopy (FS) is the use of a flexible endoscope to conduct a structural examination of the distal colon (64). The goal of FS is to examine the distal colon as closely as possible given the limitations of the endoscope length (typically 60-70cm) and the tolerance of the patient towards the procedure (65). Other endoscopes such as colonoscopes may be longer (13). FS is performed following an enema to cleanse the distal colon for bowel preparation, often using an oral laxative, which increases visibility to detect various lesions other than small (<6mm) polyps (65). One risk related to this procedure is perforation, although this is seen to be rare occurring in 1 in 25 000-50 000 cases and may actually be lower than in colonoscopies as FS patients are usually not sedated and less force is applied to colonic loops. There are additional risks linked to performing biopsy or polypectomy associated with FS in terms of bacteraemia (65). Detection rates for adenomas may differ substantially between examiners in similar patient populations. Higher detection rates have been associated with colonic distention, adequate bowel preparation, and duration of time spent in examination (13, 65). Limitations of the FS are that a negative screening test may not capture lesions which lie beyond the distal colon, and even the possibility of missing lesions within the distal colon. Risk of advanced proximal adenomas or cancer in patients undergoing FS in whom one or two small distal adenomas were found is estimated to be less than 10% which may

warrant colonoscopy based on certain patient characteristics and available resources (65). Additionally, studies have shown that the proportion of individuals with advanced proximal neoplasia, not detectable by FS, without presence of distal polyps may be as high as 46% (66). Based on these results, it is estimated that the sensitivity of FS to detect advanced neoplasms and CRC in the whole colon is 60-70% of that of colonoscopy (64).

FS detection rates in screening have been found to vary considerably. The UK Flexible Sigmoidoscopy Screening (randomized) Trial found that thirteen similarly experienced endoscopists had an adenoma detection rate ranging from 8.6% to 15.9% which was attributed to learning ability (67). Results from the Norwegian Colorectal Cancer Prevention Study employing eight endoscopists indicated a variation in detection of 36.4% to 65.5% for polyp detection and 12.7% to 21.2% for detection of any adenoma (68). Screening results from Italy also indicated detection rates of all adenomas ranging from 5.2% to 25.0% in men and 2.5%-14.0% in women with higher detection also in symptomatic individuals (69).

Other studies have examined the effectiveness of FS on various outcomes. Rigid Sigmoidoscopy (RS), using comparisons from a population of 261 individuals who had died from CRC between 1971-1988 to an age and sex matched control population on the same health plan, was shown to reduce CRC mortality. Results indicated RS in the previous 10 years prior to any cancer diagnosis showed an OR and 95% CI of 0.30 [0.19, 0.48] unadjusted and 0.41 [0.25, 0.69] adjusted for family history and frequency of periodic checkups for mortality in cancers *within reach* of the RS (70). However, adjusted and non-adjusted ORs for cancers above the reach of the RS were inconclusive being 0.96 [0.61, 1.50] and 0.80 [0.54, 1.19] respectively (70). A similarly designed case control study found that sigmoidoscopy (both rigid and flexible) decreased CRC mortality with an OR of 0.21 [0.08, 0.52] for cancer in all sites, 0.05 [0.01, 0.43] for cancer of the rectum and distal colon and 0.36 [0.11, 1.20] for cancers above the distal colon (71). The effect of FS five years prior to first diagnosis on reducing incidence of CRC was also assessed in a large case control trial of average risk individuals from the Department of Veterans Affairs in the United States. The effects of FS on colon cancer incidence indicated an OR 95% CI of 0.56 [0.46, 0.67] and 0.61 [0.49, 0.75] for rectal cancer (17).

Notably, colonoscopy was shown to have a slightly greater effect on colon cancer incidence OR 0.47 [0.37, 0.58] but a very similar effect on rectal cancer OR 0.61 [0.48, 0.77] (17). Further evidence from a case control studies indicated that the protection by FS against the incidence of distal colon cancer OR 95% CI 0.24 [0.17, 0.33] remained consistent even up to 16 years prior to cancer diagnosis (72). Inconclusive results were found for the protection of sigmoidoscopy on proximal colon cancers (72).

Randomized trials for the effect of FS on CRC incidence and mortality have also lent evidence towards its screening effectiveness. The first randomized trial for FS was the Telemark Polyp Study in Norway and was a single center study consisting of 400 randomly selected subjects aged 50-59 from the population registry in Telemark County. These cases were matched with 399 controls from the same registry who were not informed of their involvement (14). 81% of the screening group accepted FS, and were offered follow-up colonoscopy and polypectomy following polyp detection. After 13 years of follow up, CRC incidence showed a RR 95% CI for the screening group of 0.2 [0.03, 0.95] with two CRC cases in the screening and ten cases in the control group (14). A larger randomized trial also conducted in Norway from the NORCCAP trial, currently still in progress, has yielded results from the initial 7 years of follow up where population registries were used to randomize individuals aged 55-64 into once only FS screening (N=13 823) and control (N=41 913) groups from two centers (15). Of those in the screening group, half were randomized to FS and half to FS and gFOBT, with screening participation at 64.8%. Controls were not felt to have strong contamination due to the absence of an organized screening program in Norway and evidence of low opportunistic screening. Positive results involved detection of polyps >10mm and qualified participants for full colonoscopy with polypectomy. So far, after 7 years of follow-up this study has shown no indication of a significant reduction in mortality in the screening group under an intention to treat (ITT) protocol from CRC (HR 95% CI 0.73 [0.47, 1.13]) or rectosigmoid cancer (HR 0.63 [0.34, 1.18]), and no change in incidence from CRC (134.5 and 131.9 per 100 000 person years by ITT in screening and control groups). When restricting the analysis to those actually participated in screening, thereby introducing strong concerns of selection bias, total mortality was reduced for CRC (HR 0.41 [0.21, 0.82]) and rectosigmoid cancer (HR 0.24 [0.08, 0.76]). The authors plan follow-up for a

total of 15 years, and feel that the initial results after 7 years of follow-up may be insufficient to see an effect using ITT (15).

A third randomized trial was conducted in the United Kingdom across 14 centers in individuals aged 55-64 who had indicated previously they would take up offers of CRC screening, included 57 099 randomized to the screening group and 112 939 to the control group (16). The screening group was assigned to once only FS screening with polypectomy for small polyps and referral to colonoscopy if there was a positive result for being of high risk which included finding polyps >1cm, having three or more adenomas, having tubulovillous or callous histology, severe dysplasia or malignant disease, or ≥20 small hyperplastic polyps. 71% of individuals attended their screening and 5% were referred to colonoscopy with results representing a median follow-up period of 11.2 years. In intention to treat analysis, CRC incidence in the screening group was shown to be significantly reduced (HR 95% CI 0.77 [0.70, 0.84]) as was CRC-related mortality (HR 0.69 [0.59, 0.82]). Using per-protocol analysis adjusted for self-selection bias in the screening group, even stronger findings were found indicating 33% (HR 0.67 [0.60, 0.76]) decrease in CRC incidence and 43% (HR 0.57 [0.45, 0.72]) decrease in mortality. The authors express concern that their two stage selection process where initial questionnaires gauged attitudes towards screening participation may lead to a less representative sample, although they do not seem to differ so substantially from those of the NORCCAP trial which was population based (15). The Telemark study (14) is hypothesized to have a relatively larger effect estimate for intention to treat in part due to offering periodic screening services later at two and six years to those with any indication of polyps at first screening, and having the highest compliance rate (64).

C. Optical Colonoscopy

Unlike flexible sigmoidoscopy, colonoscopy permits a structural inspection of the entire colon and same-session polypectomy or biopsy of polyps and early stage cancers (13). Although initially used primarily as a follow-up intervention to other CRC screening tools, such as stool based screening tests and FS, colonoscopy use has become more widespread and use has increased substantially since 2000 in the US (73). Preparation for colonoscopy typically encompasses patients undertaking a liquid diet two days prior to

the procedure with oral laxative intake or oral lavage solutions to help clean and prepare the bowel. It is also common for some sort of mild sedative to be used during the procedure for patient comfort (13). Commonly cited disadvantages of optical colonoscopy include the time and discomfort spent in preparation for the procedure, that success is largely operator-dependent, and risk of medical complications (13). Results from a large cohort of Medicare recipients between 1991-1998 in the US have indicated that the incidence of perforation within 7 days of the procedure was 1.96 per 1000 procedures for colonoscopy and 0.88 per 1000 for patients undergoing FS, indicating an increased odds of 1.8 95% CI [1.2, 2.8] for colonoscopy (74). Trends in increased likelihood of perforation with advanced age and in presence of two or more comorbidities were also found to be significant. Perhaps most worryingly, the odds of death following a perforation from an endoscopic procedure were found to be very high for both colonoscopy related perforations at 9.0 95% CI [3.0, 27.3] and for those related to FS at 8.8 [1.6, 48.5] (74). These findings are similar to results from smaller studies (75-77). Younger individuals aged 40-59 have been estimated to have a lower incidence of perforation with a 5.2 [1.4, 19.2] fold increase in perforations in individuals 60 and older (78). Some have also hypothesized that technological improvements are reducing the risk of perforation (74). Other colonoscopy risks include various cardiopulmonary events which are typically linked to sedation, which amount to approximately half of the adverse events during colonoscopy (79).

To the best of my knowledge, there are no results from randomized trials indicating the effectiveness of optical colonoscopy on reduction of colorectal cancer incidence or mortality. However, there are at least two large scale trials currently underway in Europe to investigate this question (64). Indirect forms of evidence support the effectiveness of colonoscopy in screening as this procedure has often been used as the final diagnostic procedure for several trials investigating other screening techniques (14-17). Moreover, it is probable that the results for colonoscopy in these cases would be even stronger than the recorded effect measures due to its ability to examine the entirety of the colon and detect and remove proximal neoplasms in the absence of distal neoplasms (64). Despite this, optical colonoscopy is not always considered a gold standard screening test (13). Using virtual colonoscopy as a reference standard, evidence suggests that the miss rate of optical

colonoscopy for large adenomas (≥10mm) may be as high as 12% (80). It has also been hypothesized that polypectomy is not always successful at excising the entire polyp, leading to resulting cancers (81).

Other evidence of the protectiveness of colonoscopy has been derived from findings pertaining to CRC incidence following colonoscopic polypectomies. Data from the National Polyp Study found that those who had undergone colonoscopy with adenoma removal with follow-up colonoscopies had demonstrated reduction in the incidence of CRC in comparison to three retrospective control groups which did not receive polypectomy at highly significant reductions between 76-90% (82). An Italian study which utilized a similar case group with controls through the Italian general population found an incidence ratio of 0.34 95% CI [0.23, 0.63] (83). Other results call into question whether these findings are fully appropriate. Data from three combined chemoprevention trials for adenomas estimated post-colonoscopy/polypectomy cancer incidence of 1.74 per 1000 person years (81) compared to early estimates from the National Polyp Study of 0.6 per 1000 person years (82). This difference has been hypothesized to result from differing recruitment criteria and differences in the completeness of follow-up (81).

Ecological level data has similarly been used to establish the effectiveness of colonoscopy. Data from the US has shown that, upon colonoscopy being included in Medicare reimbursement for the purposes of screening for high risk individuals in 1998 and all individuals in 2001, this was associated with much greater use of colonoscopy and associated with significantly increased diagnosis for early stage cancers from 22.5% in the initial period, to 25.5% after 1998, and 26.3% after 2001 (84). The authors posit that much of this difference seemed to be driven by detection of proximal lesions, derived from whole-colon screening offered by colonoscopy (84). In Canada, using a population based fixed cohort of individuals aged 50-90 followed from 1993-2006, data from public insurance and services sources for Ontario was used to examine the relationship between the rate of colonoscopy and death by CRC (85). Using a Cox multivariate model controlling for age, sex, comorbidity, income and residence, a hazard ratio of 0.97 95% CI [0.95, 0.99] indicating that for every 1% increase in colonoscopy in the region of the cohort member was associated with a 3% reduction in hazard of death by CRC (85).

D. Effectiveness of Colorectal Cancer Screening using gFOBT with Follow-Up

Several randomized trials have lent evidence towards the effectiveness of colorectal cancer screening regimens initiated by annual or biennial fecal occult blood test with follow-up endoscopy. The success of these programs is determined by the use of several screening tools which vary both by study and by time. In several cases these studies lend evidence to the effectiveness rather than efficacy of an organized CRC screening program as intention to treat (sometimes referred to as intention to screen in the literature) conditions are often used and most of the randomized trials make use of a population sample, offering screening services through the mail rather than recruiting volunteers.

Some case control studies have examined the influence of screening using fecal occult blood tests. These studies share limitations of selection bias and thus the results should be interpreted with caution. One such study took a population of clients in the Kaiser Permanente medical care program and compared 485 individuals who had developed fatal CRC to sex and age matched controls (86). Screening gFOBT with follow-up controlled for personal and family history of CRC and frequency of healthcare checkups was associated with an OR of 0.69 95% CI [0.52, 0.91] in five years prior to case diagnosis (86). Results of a similarly designed trial in Germany for screening up to three years prior to diagnosis indicated that males neither had high screening rates for gFOBT or a significant protective effect demonstrated OR 0.92 [0.61, 1.75] although such a protective effect was observed in females OR 0.43 [0.27, 0.68] who were also noted to have undergone screening more frequently than males (87).

In the UK, 152 850 individuals aged 50-74 living in the Nottingham area between 1981-1991 were recruited from the population and randomly allocated to receive gFOBT screening or no screening (36). This study was designed as to produce a study sample which was not self-selected, and those who were controls had no direct knowledge of their participation in the study and the screening group was sent a Haemoccult gFOBT with instructions from their physician without pre-warning. Subsequent screening was offered to those with a negative gFOBT or follow-up colonoscopy every two years and the median length of follow up was 7.8 years. With 38.2% of participants completing all FOB tests, 21.4% completing at least one test, and 40.4% not completing any test, the ITT

rate ratio for deaths from verified CRC was found to be 0.85 95% CI [0.74, 0.98] indicating a decrease in the screened group which was found to be even stronger for the subset of screened individuals who complied initially and took the first screening test 0.61 95% CI [0.50, 0.74]. There was found to be a non-significant difference in CRC incidence with a rate ratio of 1.04 [0.95, 1.14] (36). At a median follow-up of 11 years up to 1999 past the expiration of the screening program in 1995, there was a rate ratio of 0.73 95% CI [0.57, 0.90] for deaths from verified CRC, and a rate ratio of 0.98 [0.83, 1.13] for CRC incidence and no evidence for effect modification by sex or site of cancer (88).

In Denmark, a similarly designed population based randomized trial was performed in which 30 967 people were selected in the city of Funen aged 45-75 for invitation to screening with gFOBT and 30 966 were uninformed controls. The primary difference to Nottingham was that only individuals who had responded to first screening invitation which had 67% compliance were invited for four subsequent screening rounds, each of which had compliance above 90%, although non-compliers remained in the sample. After 10 years of follow up stating in 1985 there was a significant difference observed in the screening group in CRC mortality reduction with mortality ratio of 0.82 95% CI [0.68, 0.99] (38). After 17 years of follow up in which time a total of nine biennial screening rounds had been offered, those in the screening group had a reduction in CRC mortality with a mortality ratio of 0.84 [0.73, 0.96] although the mortality ratio including deaths from complications associated with CRC treatment was 0.89 [0.78, 1.01] and no significant results were found towards preventing CRC incidence despite removal of 2-4 times as many large adenomas (89). A further study of this type in Goteburg, Sweden, randomized subjects aged 60-64 at study entry and after follow-up ranging from 2-7 years with 63% participation in the initial screening and 60% participation in rescreening two years later. This study did not have sufficient follow-up time to establish any trend in CRC related mortality, although the trial established that significantly more early stage carcinomas were being diagnosed in the intervention group and significantly fewer late stage carcinomas were being diagnosed in the intervention group (37).

In the US, the Minnesota Colon Cancer Control Study took a slightly different approach by randomizing a group of volunteers rather than population groups to annual or biennial

screening or control, with gFOBT with various follow-up procedures as the screening tools. After 13 years of follow up, participation was 75.2% and 78.4% in the annual and biennial screening groups respectively, and, like other studies, there was no observed effect of screening on CRC incidence there was an observed protective effect found in the annual screening group with a mortality ratio of 0.67 95% CI [0.50, 0.87] but not observed in the biennial screening group 0.94 [0.68, 1.31] (90).When this trial was studied for 18 years of follow up cumulative incidence ratios indicated reduction in CRC incidence in the annual (0.80 [0.70, 0.90]) and biennial (0.83 [0.73, 0.94]) screening groups compared to controls (58) and additionally, CRC related mortality was found to be reduced in the annual (rate ratio 0.67 [0.51, 0.83]) and biennial (0.79 [0.62, 0.97]) screening groups compared to controls (91). The voluntary nature of this study may have resulted in an atypical sample as higher participation rates were observed in this study as opposed to others.

A relatively recent Cochrane systematic review of screening used data from the previous four randomized trials to establish an overall reduction in relative risk of CRC mortality from screening of 0.84 95% CI [0.78, 0.90], with a RR of 0.85 [0.78, 0.92] when looking at biennial screening results, and when adjusted for attendance to at least one round of screening, the predicted relative mortality reduction was estimated at 0.75 [0.66, 0.84] (18). Other population based evidence from France has also contributed to the knowledge of CRC screening. Using prospective non-randomized methods, 12 administrative districts in Burgundy were offered biennial gFOBT screening with 17 similarly sized administrative districts as controls. The study included 11 years of follow up with 69.5% of those invited participating in at least one screening round and 38.1% adhering to all screening rounds, and resulted in a non-significant CRC incidence ratio of 1.01 95% CI [0.91, 1.12] and a significant CRC-mortality ratio of 0.84 [0.71, 0.99]. Additionally, when departing to a per protocol style analysis to investigate the reduction in CRC mortality among those who screened at least once in the intervention group, the mortality ratio showed a stronger reduction in the screened group of 0.67 [0.56, 0.81] (92).

E. Screening in Canada

Prior to the implementation of screening guidelines in the early 2000s, screening in Canada was very low[2]. Using Ontario as an example, the proportion of individuals aged 50-74 in 1992 who reported undergoing a gFOBT was approximately 6%. In 1999 this number in individuals aged 50-69 increased to 15% for testing in the previous two years and by 2001 there was a noticeable 20-30% increase in gFOBT screening as estimated from OHIP billings (20). Data from clustered random samples of patients from Ontario hospitals in 2000 found that, in those newly diagnosed of CRC, only 6% of diagnoses were through screening. Additionally, among those with a discernible cancer stage, 43.4% of those with newly diagnosed cancer had an advanced stage of cancer indicating diminished chances of survival (93). The authors of this study also examined colonoscopies and found that approximately one quarter of colonoscopies performed were for purposes of screening, with the majority being performed due to the presence of symptoms (93). A 2004 survey associated with the Ontario FOBT Project in Ontarians 50 and older found that half of the respondents had not heard of gFOBT screening and 22% of those surveyed were not considering screening (20). Results from consecutive colonoscopy patients in a single Vancouver based centre in 2000-2002 indicate that a large proportion (65%) reported undergoing colonoscopy due to symptoms and the remainder used the procedure for screening (94). Screening colonoscopy was also found to be more common in men and in those with higher incomes and education levels (94).

As mentioned above, several organizations within Canada have made recommendations for screening for colorectal cancer several years prior to the implementation of any provincial organized colorectal cancer screening program. In 2001, the Canadian Task Force on Preventive Health Care had recommended annual or biennial FOBT in periodic health examinations for individuals older than 50 with FS as an additional screening option (7). In 2002 similar recommendations were released from the National Committee on Colorectal Cancer Screening which advocated for biennial screening for individuals 50-74, with follow up of positive results by colonoscopy or FS or barium enema if

[2] Throughout this project, unless explicitly stated otherwise, I will assume that all mention of fecal occult blood tests in the literature or in the CCHS are in regard to guaiac fecal occult blood tests and will label them as such even if this was not specified in the original source.

appropriate (6). In 2004 the Canadian Association of Gastroenterology and the Canadian Digestive Health Foundation added to this by recommending biennial FOBT for individuals 50 and older with follow up colonoscopy, or FS or barium enema every five years or colonoscopy every ten years (8). It is important to note that these screening recommendations were not yet accompanied by any form of organized colorectal cancer screening program although all of these services were available to patients.

Evidence indicates that, despite the presence of screening guidelines, CRC screening remained suboptimal. Data from the 2003 Canadian Community Health Survey (CCHS) on the provinces of Newfoundland and Labrador, Ontario, Saskatchewan and British Columbia, using an average risk population of individuals aged 50 and above without past or present history of CRC, indicated that adherence to the recently implemented screening guidelines was low (19). Adherence for gFOBT screening guidelines, defined as gFOBT within previous two years, were only 15.1% and adherence for combined gFOBT and endoscopy defined as FOBT in the previous two years or colonoscopy or FS within the previous 10 years, were 30.1%. Other results found that many individuals who were undergoing gFOBT or endoscopy were not doing so according to recommendations (19).

However, there is strong evidence to suggest that in many provinces there has been a general trend towards increasing proportions screened in the average risk population. Data from the 2008 Canadian Community Health Survey has indicated that screening adherence, redefined as gFOBT in the previous two years or flexible sigmoidoscopy or colonoscopy in the previous five years, had risen to a national average of 40% in those aged 50 and over (23). CRC screening data indicated that, in this population group, proportions screened had increased significantly between 2005 to 2008 in Newfoundland and Labrador (26.6-33.9%), New Brunswick (27.6-34.5%), Ontario (37.9-49.6%), and from 2003 to 2008 in British Columbia (26.6-36.8%) although information on all provinces was only available in the CCHS 2008 and it is possible that other provinces have also undergone significant increases in screening in the average risk population. These increases were significant for increases in both gFOBT and endoscopy tests in all these provinces. These increases are thought to be, in part due to a greater awareness and use of screening guidelines over time and the highest screening adherence which was

observed in Ontario and Manitoba were thought to be in part due to the effects of organized screening programs there although this was not directly tested (23).

Contacting physicians directly has been used as a way to ascertain the degree to which physicians report recommending screening and on what target groups they are focusing. Data from the FOBT pilot project in Ontario during 2003-5 has indicated that physicians were often confused about screening guidelines and believed that discussing the topic with patients was overly time consuming (20). A survey of recommendations for screening behaviours of a number of family and specialist physicians in Calgary, Alberta, was administered at approximately the same time. With a 61% response rate, 58% of physicians reported recommending screening to healthy individuals without family history of CRC, and 96% reported recommendations to those individuals with a family history. Patients of 50 years of age were identified as the group most likely to start receiving screening recommendations (64%) with those at age 40 being the second highest. gFOBT was the most common initial test in the average risk group (79%) and colonoscopy was the reported most common initial test for the advanced risk group with family history (56%) (21). Information on CRC screening has also been asked of physicians themselves from a survey sent to members 50 years of age and older from various physician specialist groups across Canada. Screening rates were seen to vary by specialty with radiologists and gastroenterologists reporting the highest screening adherence at 61% and gynecologists and obstetricians screening at 44%. Physician screening was highest in Ontario and lowest in the Maritimes, primary screening was most often colonoscopy (56%) and secondly gFOBT (27%), screening was less common in those aged 65 and older, and screening modality preference varied by province and specialty. Non-screened physicians most commonly reported they intended to screen in the future but lacked sufficient time (49%) and in some cases reported doubts that there was sufficient evidence for screening (34%) (95).

Another technique which has been employed in Canada to measure screening has been use of a regression discontinuity design to see if there is evidence of significant increases in screening in individuals who, at age 50, enter the recommended age for screening compared to a very similar group with age <50. Using CCHS from 2003 and 2005,

proportion screened, defined as having had gFOBT or endoscopy in the past two years in the average risk population across Canada, rose from 15% at age 49 to 17% at age 50 and 21% at age 51 (96). With an age range of individuals from 40-60, evidence for significantly increased proportion screened in the average risk population upon entering the threshold age of 50 was not found in Atlantic provinces, Ontario or British Columbia although was found for Saskatchewan adjusted OR 95% CI 3.62 [1.09, 12.05]. This could also be represented as a 1.3 percentage point increase in the probability of being screened across Canada with Saskatchewan indicating a 12.5% increase (96). Notably, similar models on adherence to screening guidelines for breast and prostate cancers were typically not observed to have been met in addition to that of CRC screening (96).

F. Determinants of Screening

Screening adherence has been demonstrated to differ between socioeconomic groups. Throughout the literature, some studies indicate that women are slightly more likely to complete gFOBT tests than men, although this varies between studies and in different study contexts (97, 98). Age is a somewhat strong and consistent predictor of screening behavior across many studies for gFOBT although this relationship seems to decrease in advanced ages, such as those >80 years of age, and is does not display significant trends for FS (98). Studies which examine marital status and screening have provided conflicting results (97). Additionally there have been significant, although not always consistent, positive relationships observed between socioeconomic variables such as income and education for gFOBT (97, 98) including SES measured at the area level (99). One study of the organized CRC screening program in the UK examining screening in the city of London found great disparities based on area socioeconomic deprivation where FOB test completion was 49% in the least deprived quintile and decreased to 32% in the most deprived quintile. The authors noted that, controlling for ethnic diversity, household mobility and health status, SES area level deprivation accounted for 62% of variance in return rates (100). Factors related to insurance coverage and affordability have been indicated to influence CRC screening and choice of screening procedure in the US (101) which lends evidence as to why even insured individuals may seek limited screening options. Out of 180 private insurance health plans in the US of varying types, 97%

offered FOBT, 97% FS and 88% colonoscopy for non-Medicare enrollees. However, in some cases coverage was limited to those patients who were identified as high risk. Additionally, various forms of cost-sharing were common, ranging from an average of 30% for FS and 33% for colonoscopy. It is also notable that approximately only two thirds of insurance plans interviewed noted that they had implemented guidelines for CRC screening, and that in many cases guidelines were incomplete and missed instruction on how to follow up with positive FOBT or FS (101).

Physician support and advocacy can be highly influential in adherence to screening. An Australian randomized trial with three groups of patients consulting general practice physicians found significant differences in the ways in which screening was endorsed. Screening participation from a group which just received an invitation letter without indication of their physician's involvement was 32%, participation was 38% when a letter from the physician practice was received, and when this letter included the personal signature and letterhead of a practice physician the participation was 41% (40). An Italian RCT found that physicians who worked in a general practice setting were much more likely to persuade patients to finish screening tests at a RR of 3.4 95% CI [3.03, 3.70]. This trial also established that physicians who see more than 26 patients a day have lower compliance rates, possibly indicative of a lack of communication, and that physicians who were unclear on screening guidelines had lower compliance rates (102). Further evidence supporting physician involvement is found from a prospective trial in Israel where patients eligible for gFOBT were sent either a reminder for screening by letter or by phone, or received no reminder. Screening adherence increased significantly comparing the reminder groups to the control with telephone communications found to be more effective than sending a letter (103).

The evidence for greater adherence with physician involvement coincides with the idea that perceived value of the screening test is an important predictor of screening. Several studies have found that beliefs that screening is efficacious and that polyp removal can prevent colorectal cancer, which is perceived as a risk, are important predictors of intention to screen and adherence (97, 104-106) although simply perceiving CRC vulnerability is not consistently found to be a predictor of screening (98). This

observation may have links to patient education and the trustworthiness and credibility of physicians, and their subsequent influence on patient behavior (40). It might also relate to another observation that healthcare seeking behaviours or preventive behaviours indicated by such practices as regular medical or dental check-ups are very consistent predictors of screening adherence in the literature (97). Having had a previous gFOBT, for instance, is a fairly consistent predictor of undertaking gFOBT later although this is not consistently the case with FS (98).

Certain structural factors can influence screening. For instance, likelihood of compliance has been tied to such factors as geographical accessibility of screening services and time required for screening (102, 107). Embarrassment and feelings of anxiety towards screening have been significantly associated with likelihood of not completing gFOBT, even amongst those who believe in its effectiveness (98). Results from the UK screening program indicate that individuals with an ethnic background from the Indian subcontinent are less likely to complete the FOB test, possibly indicative of cultural difficulties pertaining to belief of the unhygienic and distasteful nature of the test (100).

Some effort has been made to examine various factors which coincide with SES and may influence screening behaviour. Interviews associated with participation from the UK screening program found that although factors such as stress and social support were associated with low SES individuals, these did not ultimately have much influence on screening interest (108). Low SES individuals were more fearful of cancer, more fatalistic about cancer and less likely to hold beliefs that lifestyle change would create health benefits. So, whilst low SES individuals may not be underestimating their risk of CRC or be inhibited by cost, little value is perceived in screening or lifestyle modification (108), which has previously been indicated quite consistently to anticipate screening behavior (97). Additionally, there is evidence of disparities for counseling itself from the US population which finds that individuals who are less educated or non-White are less likely to report receiving screening recommendations or counseling from their physicians (109, 110) which relates to both knowledge of and holding value in screening.

Results from screening adherence patterns in Canada share many of the characteristics seen above. Results from 2003, prior to the implementation of any organized CRC

screening program, indicate that a having a high as opposed to low household income status was associated with higher screening guideline adherence as was having a full or part time job (gFOBT in past two years or endoscopy in past 10 years) (19). Evidence also suggested that those who were health conscious or showed evidence of following a healthy lifestyle were more likely to adhere to the screening guidelines as being moderately or highly physically active, being a non or former smoker, and reporting having taken a flu shot were all significant predictors of screening. Having a regular physician was also highly predictive of screening OR 95% CI 2.39 [1.87, 3.03] (19). 2008 data largely correspond to these findings where those who were in the lowest income quintiles, those who were daily smokers, had BMI≥40, and those who were physically inactive were found to have significantly reduced likelihoods of screening adherence. Notably, in many instances such as those who self-identify as obese class I and II and those who are occasional smokers, moderately active, and report being in any household income quintile but the lowest there are not significant differences in screening, possibly indicative that screening is not being primarily undertaken by those with the highest health consciousness or the highest socioeconomic positions. Additionally, further results indicate that having a regular physician was a highly significant predictor of screening adherence with adjusted OR 3.0 [2.1, 4.4] without visiting in the previous year and both having a regular physician and reporting contact with the physician in the previous year increases the OR to 5.7 [4.0, 8.1] and even simply visiting a physician in the previous year, even in absence of a regular physician, was found to be a significant predictor of screening at OR 2.4 [1.5, 3.7] compared to not having a regular physician and not visiting in the previous year (23).

Ontario-specific results on reasons for screening non-adherence have been collected from qualitative interviews circa 2007 from a convenience sample of Ontarians aged 50-90 (111). Reasons for declining CRC screening included categories related to having a lack of symptoms, concern over the painfulness and scariness of colonoscopy, belief in efforts to engage in healthy lifestyle routines, such as exercise and healthy diets, would substitute for CRC screening, and deliberate decisions not to undergo screening due to advanced age especially with invasive treatments. Additionally, 24% of this sample expressed ignorance of CRC incidence and mortality rates and education was seen with some

consistency as a reason for this lack of knowledge (111). A recent telephone based survey on Ontarians 50 and older has found that out of 54% of women and 45% of men had heard of FOB tests and this familiarity was noted to be increased in individuals with higher education levels and in married or equivalent relationships (112). A choice-format stated reference survey design was also conducted in southern Ontario on a random selection of patients aged 40-60 from primary care networks to pinpoint patient preferences for screening (113). One result of this survey was that preference for no screening was as high as 30%. The order of the importance of test attributes was highest for test accuracy, then test preparation, test process and finally pain. As such, overall, patient preferences were most in line with that of a CT scan virtual colonoscopy with limited-invasiveness, minimal preparation, lack of pain, and high accuracy and lowest for gFOBT, largely due to poor accuracy. Individual features of those who were older than the median age of 51, were female, had less than a college education, higher than median income, in poor health, and had a family history were more likely to state screening preferences than no screening (113). These results help reinforce the importance of the recommendation for screening by a physician in influencing average risk individuals to undertake screening for CRC.

G. Summary of Findings

This section has reviewed the research findings on the guaiac fecal occult blood test, flexible sigmoidoscopy, and colonoscopy. Evidence for the effectiveness of the gFOBT test is present although there is a tremendous degree of variation in the reported effectiveness of this test given different tests, modes of preparation, and study designs. Results from the highest quality studies indicate it is likely that the sensitivity of the gFOBT on single application is approximately 30% and specificity between 80-90% for detection of CRC. There is also evidence from randomized trials for the effectiveness of FS and indirect evidence for the effectiveness of colonoscopy in reducing CRC incidence and mortality. Additionally, several large population-based randomized trials testing an organized screening strategy with gFOBT as the initial screening test have been undertaken and shown significant effects in reducing CRC mortality. Screening adherence in Canada has historically been lower than preferred despite the risk of CRC.

Predictors of screening adherence tend to vary to some extent but with some consistency indicate that advanced age, physician support and participation, and some measures of socioeconomic status influence likelihood of screening.

There are some notable gaps in the literature in regards to an organized screening program. Although information on screening practices in Ontario over a time period in which an organized screening program was underway has been published, it is not clear as to how much of this effect can be attributed to this program. Although it was noted that screening in Ontario in 2008 was significantly increased from earlier (23) this does not indicate a positive effect derived from the program because several other provinces also experienced a similar increase. Without accounting for general trends towards greater screening adherence elsewhere it is unclear from examining these findings in proportions screened what the cause may be. Additionally, more recent information would improve on these findings to establish if this increase was maintained. A further detail is that by examining multiple year adherence measures (2 for FOBT and 5-10 for endoscopy) there is much less certainty in regard to attributing screening outcomes to an intervention initiated at a specific time. Evidence of the RCTs conducted in the UK, US, Sweden and Denmark is promising to justify the realization of a screening program but hardly serves as a measure of proof for the effectiveness of one. The experimental designs of these studies do not reflect a real world application of a screening program where participants may act differently under experimental conditions.

Section III: *ColonCancerCheck*

A. Background

Prior to the implementation of *ColonCancerCheck*, screening for colorectal cancer in Ontario was accomplished through opportunistic screening of at-risk individuals. Opportunistic screening relied on incentives based in fee-for service payments to physicians, although this gave rise to concerns about inappropriate screening and lack of quality assurance measures (22). An organized screening program which could provide

consistent quality-assured screening methods to target specific risk groups, with both management arms for policy implementation and medical arms for healthcare treatment and decision-making, was seen as a way to solve the difficulties integral to opportunistic screening (20, 22). Other actions to confront deficiencies in provision of colorectal cancer screening called for establishment of standards and recommendations for the procedure and in 2001 the Canadian Task Force for Preventive Healthcare made the first of this kind of recommendation in Canada and endorsed colorectal cancer screening with guaiac fecal occult blood test (gFOBT) or with periodic flexible sigmoidoscopy in average risk individuals (7).

A pilot study was developed in a partnership between Cancer Care Ontario (CCO), the Ontario Ministry of Health and Long-Term Care (MOHLTC), the Institute for Clinical and Evaluative Sciences (ICES) and the Ontario Association of Medical Laboratories (OAML) to investigate opportunistic screening for colorectal cancer in Ontario in 2003-5 (22). This study was conducted on volunteers aged 50-75 in 12 of 37 public health regions in Ontario which were randomly selected and seen to be representative of Ontario as a whole based on geographical composition and availability of healthcare services (114). Printed health education materials on colorectal cancer (CRC) and gFOBT were provided to participating primary healthcare providers (PCPs) and patients in six health regions which served as a 'primary care arm'. A more open approach to intervention was left up to public health units in the other six intervention regions which included general mass media and community level promotions in addition to direct advocacy of physicians, and finding physicians for participants without a PCP. Additionally, a laboratory requisition and consent form were supplied, and a gFOBT kit was provided on condition that the consent form was filled out and the kit returned to the laboratory for analysis. Consent was also provided for researchers to have access to participant's Ontario Health Insurance Numbers (OHINs) to observe health procedures taken and for data linkage. A total of 6972 participants filled out consent forms and handed in gFOBT kits. 96% had negative results, with 2.8% being positive and 1.2% being inconclusive (114). The authors observed that colonoscopies and polypectomies were more frequent in study participants than in the rest of the population, and that 64.6% of those with a positive gFOBT chose to have a colonoscopy by the end of the follow-up period. They also

observed that men were more likely than women to have a follow-up colonoscopy (73% vs. 56%) and had a lower median time between having a positive gFOBT and colonoscopy (121 days vs. 202 days) (114).

Some evidence suggested that screening adherence in public health units was slightly higher than screening through PCPs. Also, several interventions were observed to increase screening including mass media promotions, detailing PCPs, and improving access to gFOBT kits (20). The authors took the length of the times from positive gFOBT to endoscopy follow-up as an indication that services where not operating at sufficient capacity to provide rapid screening. A survey associated with the study found that 67% of PCPs reported following through on recommendations of gFOBT to age eligible patients, although 45% felt that explaining the procedure to patients took up too much time and nearly half felt that gFOBT had too many false results to be a useful screening tool. Many physicians were also noted in interviews to be confused on issues concerning the appropriate ages to screen and means of screening (20).

There was also some indication that, despite low proportions of individuals screened, the pilot program lead to increases in those regions screened compared to the public health regions in which there was no pilot program. Comparing 12 months prior to the program to 12 months into the program, an 18% increase in total gFOBT screening occurred in health regions without the pilot whereas a 42.7% and 41.5% increase in screening was observed in the six primary care and six public health regions participating in the pilot, respectively. Even those who were not registered participants living in participating health regions showed a very similar increase of 38.3% (20). This was also some evidence for a general increase in CRC screening independent of the pilot program. Insufficient data on positive gFOBTs was collected to make many inferences between subgroups of patients. Based on the results of this pilot trial, Cancer Care Ontario recommended the implementation of an organized CRC screening program (20) which were as follows.

1. The Ministry of Health and Long Term Care Establish a Provincial Population Based organized CRC screening program.

2. The program should be phased in over a four year period based on meeting predetermined program goals of recruitment and follow-up, and continue to roll out after these success factors have been achieved.
3. Cancer Care Ontario should implement a mass media awareness campaign for both PCPs and age eligible residents of Ontario aged 50 and older in addition to individual health promotion messages.
4. Establishment of a central program office at Cancer Care Ontario for the screening program, planning and implementation of the program, contracting with laboratories to ensure consistent high quality gFOBTs, and contract with hospital based endoscopists to for high quality timely access to endoscopy. This office will additionally implement an information system to monitor quality and success factors of the program. The program structure is recommended to be able to accommodate a change in screening method(s) at either the level of stool assay or endoscopy in the future.
5. Automatic updating of a population list of persons eligible for screening.
6. CCO and MOHLTC should meet with PCPs to ensure needs and concerns are met and to establish a collaborative relationship.
7. Bids should be taken from laboratories for services involving sending invitations and letters to individuals in Ontario for screening, analyzing kits, and sending the results to the appropriate parties.
8. Establishment of regional offices for telephone assistance, and assisting individuals with positive results.
9. Critical success factors of the program to be monitored:
    - 65% adherence in age eligible population
    - 75% of positive gFOBT results followed up with colonoscopy within 12 weeks
    - 60% of invasive cancers detected at stage I
    - 95% advanced non-invasive neoplasms to be resected by polypectomy
    - Colonoscopy related complications should be lower than 3/1000 for bleeding, 1/1000 for perforation and 1/15 000 for death.

B. Development of *ColonCancerCheck*

Funding for *ColonCancerCheck* was first announced in January 2007 by the Ministry of Health And Long Term Care in Ontario in cooperation with CCO (22). The initial program features were also outlined at this time. Such features included the distinction between average and advanced risk individuals, biannual screening for those of average risk by gFOBT, and colonoscopy for those of advanced risk (22). Other important features included the recommendation of central roles for PCPs, provisions for individuals without PCPs to be able to obtain a kit through pharmacies or through calling 1-800 numbers, use of a single gFOBT kit to ensure quality and consistency, registry creation for the purposes of sending invitations, result and recall letters, ongoing program evaluation, and a five year campaign to help educate the public and healthcare providers on such things as the benefits of CRC screening and early detection (22).

In the subsequent 15 months, efforts were directed at planning and development of the program. CCO undertook literature searches to develop standards and guidelines related to physician administered colonoscopies and physician training, institutional standards for patient assessment, infection control, monitoring during and after conscious sedation, resuscitation capabilities, and performance standards related to bowel preparation, perforation rates, cecal intubation, sedation and pathologies (115). Endoscopists administering colonoscopies would agree to perform a minimum of 200 colonoscopies a year and perforation rates would need to be lower than 1 per 1000 in all colonoscopies and 1 per 2000 for screening colonoscopies (115). Similarly, standards for gFOBT were made based on the performance of the kit, the accommodative aspects and usability of the kit, and laboratory procedures for processing the information (53). These recommendations called for a single gFOBT kit with sensitivity $\geq$40% and specificity $\geq$95% to detect CRC in repeated testing. Decisions were made that a positive result would be indicated by having one or more positive windows on the gFOBT card and indeterminate results would incur when there were one or more windows with uncertain results with no positive windows (53). The kit was to make minimal restrictions on diet and medication intake, and was to be made by a manufacturer who would work with the organization to help ensure that the kit was easy to use, and would allow participants to safely mail their

samples in leak-proof envelopes using Canada Post Standards. Recommendations also followed that laboratories would meet certain quality standards and process a sufficient quantity of kits to become proficient in the procedure, develop quality control protocols, and participate in the monitoring of results. Recommendations in regards to information collection, processing of incomplete samples, time and following manufacture instructions were also made (53).

To address the capacity of hospitals to meet an expected increase in colonoscopy procedures as a result of the program, additional funding was allocated to public hospitals in April 2007 (22). This funding was contingent on the agreement from hospitals to follow the CCO's recommendations for colonoscopy standards and for hospitals to provide CCO with data pertaining to colonoscopies which included information on number and type of procedures, wait times, and performance issues (22). The Ontario Association of Medical Laboratories (OAML) made agreements with the MOHLTC to adopt gFOBT laboratory standards from CCO and implement changes to collect gFOBT data for CCO (22). Regulatory changes were made so that nurse practitioners would be able to order gFOBT kits for their patients. Prior to the launch of *ColonCancerCheck* gFOBT kits were sent out to over 10 000 physician and nurse practitioner offices and 3000 pharmacies. Assessable instructions for patients pertaining to FOBT were translated into 27 languages. A public awareness campaign was developed to provide information on CRC and screening to the general public through television advertisements, websites, posters, pamphlets, and street teams who administered public literature at public events (22). An education campaign was also started for healthcare providers prior to the launch of *ColonCancerCheck* and included patient information kits, counseling manuals, medical education events and regional forums (22). Asymptomatic average risk individuals were defined as those aged 50 or greater with no bowel symptoms and no family history of CRC (116).

C. *ColonCancerCheck* Program Features

ColonCancerCheck was formally launched by the Minister of Health and Long-Term Care on Mar 14, 2008. Separate recommendations were made to those of average risk, where gFOBT was recommended, and those of advanced risk, where colonoscopy was

the recommended first test. A single make of gFOBT kit was used by all laboratories to help ensure quality, consistency and interpretability of results, and was administered with postage-paid return envelopes and assessable instructions (22). Requirements were fairly modest and stated that the subject should not ingest vitamin C or citrus juice or fruit three days prior to stool sampling. In October of 2008, CCO launched the Primary Care Strategy to increase the level of CRC screening through engaging with PCPs (22). A quality determinants framework for the purposes of program evaluation was chosen using the European Union guidelines for colorectal cancer screening, specifically measures for evaluating and interpreting screening outcomes (117).

The gFOBT used by *ColonCancerCheck* has a reported sensitivity of 13%-25% on the first test and 50% upon repeated testing for detection of colorectal cancer. The specificity of the gFOBT for CRC detection is 80%-90% on a single test and 96%-98% after repeated testing (116). The kit brand and manufacturer is "Hema Screen; Immunostics, USA" (118) which is also used in the United Kingdom's organized CRC screening program (119). Results from the UK pilot for the screening program in a population of 478 250 residents aged 50-69 indicate a positive test result rate of 1.9% and a positive predictive value of 10.9% for CRC and 35.0% for adenomas. Slight differences were observed between men and women and for positive predictive values of cancer detection as well as between England and Scotland, although more substantial differences were found in the PPV for all neoplasia (cancers and adenomas) detection in men (England 53.7) compared to women (England 36.1) (119).

A positive test will result in the PCP being informed and arranging a colonoscopy for the patient via a dedicated fax. A negative result would result in a letter to the participant and recall in two years from *ColonCancerCheck* (116). Patients are able to obtain a gFOBT kit from a pharmacy only if they meet the criteria of being average risk and are unattached, that is, they have no PCP (120). A gFOBT requisition form is used to determine program eligibility and is available at pharmacies. If gFOBT results are unreadable or incomplete, a letter is sent to the participant instructing them to return to a pharmacy or Telehealth Ontario to obtain another kit. If the test is negative, the participant is sent a letter indicating this and will be contacted again in two years' time. If

the test is positive, *ColonCancerCheck* will facilitate an appointment with a family physician for follow-up care in which a colonoscopy may be scheduled (120). In future years, *ColonCancerCheck* aims to continue to further implement public and PCP directed education and awareness materials and to continue to monitor program quality measures and outcomes. Further initiatives are to develop strategies for access to care for certain populations such as members of the First Nations community and create pilot studies to test screening adherence through invitation based screening initiatives (22).

D. Program Performance: Evidence Thus Far

There has been some initial evidence for the success of *ColonCancerCheck* to increase screening rates in Ontario. Using data from the OHIP Claims History Database for colonoscopy and gFOBT service history and the Registered Persons Database (RPDB) for demographic and geographical information from the MOHLTC Cancer Care Ontario, it was found that 29.7% of Ontarians 50-74 had a record of at least one complete gFOBT in the previous two years in the period 2007-8 (inclusive) which was an increase from 19.9% in 2005-6 and approximately double that of 14.8% in 2003-4 (22). Participation was noted to be slightly higher in women (31.1%) than men (28.8%) although this difference was not seen in older men. Participation was highest in those aged 65-69 (34.6% in 2007-8) and lowest in those aged 50-54 (24.5% in 2007-8) and varied substantially by health region. 62.1% of individuals with a positive gFOBT had a follow-up colonoscopy within six months with no difference between men and women and only minor differences by age. Data was collected on the gFOBT-positive rate which, although not very helpful in indicating the predictiveness of the test, showed that men overall had a higher gFOBT-positive rate than women (5.3% vs. 3.5%) but did not seem to vary substantially by age. Program invasive cancer detection rate, proportions of cancers detected within six months of gFOBT or colonoscopy in those with family history, was 2.3 per 1000 overall with the rate being higher in men at 3.1 per 1000 than women at 1.7 per 1000 and was highest amongst the oldest age group at 70-74 years of age. Program invasive-cancer detection rate for gFOBT only was 2.3 per 1000 in men and 1.2 per 1000 in women (22).

E. Organized CRC Screening Programs Elsewhere

In the province of Manitoba, phase 1 of a colorectal cancer screening program was conducted in April 2007 to September 2009 and was limited to the regional health authorities of Winnipeg and Assiniboine (121). The pilot used Haemoccult II Sensa as a gFOBT kit and targeted average risk individuals aged 50-74. Overall participation was noted to be 18.1% and initial results estimate a positive predictive value from gFOBT of 16.7% for advanced adenomas and 8.3% for CRC, although these numbers are only based on 30 results combined of both these outcomes. The program, *ColonCheck Manitoba*, will continue to expand to offer services through direct mailing province-wide (121).

The province of Alberta shares many of the aspects of a screening program in Ontario including gFOBT kits obtained from primary care providers and sent to central labs, providing educational materials to at risk individuals and healthcare providers, providing a follow-up reminder system for healthcare providers, quality assurance measures, and quality improvement measures. Like both Ontario and Manitoba, this program, the *Alberta Colorectal Cancer Screening Program*, targets average risk individuals and like Manitoba limits those individuals to the age range of 50-74 (122). This program is currently under development with initiatives from working groups for such activities as recommending screening protocols and clinical criteria, investigation of strategies for the services delivery model, providing further health education materials and awareness strategies and development of a performance monitoring framework (123). To the best of my knowledge, although a more detailed business plan was to be released in Fall 2010 (123) none has been made available.

F. Summary of Findings

ColonCancerCheck was developed in order to shift the common practice from individualistic and opportunistic screening to a more systematic method with greater levels of adherence. Greater amounts of funding were made available to hospitals to meet the anticipated increase in demand for colonoscopies and a number of standards for care pertaining to gFOBT and endoscopy provision were made. Publicity and educational materials, the development of a registry for invitations and result letters, and avenues for

individuals without PCPs to obtain gFOBT kits were also part of this intervention. There is some evidence from OHIP Claims History Database which established that gFOBT screening had increased substantially in Ontario in 2007-8 period from the 2005-6 period although it is not clear how much of this increase is due to the program and how much might be due to a broader trend towards more screening (22). There is also some evidence of changes in screening practice from the pilot. This pilot program found that, in comparing screening 12 months before and after introduction of the pilot, gFOBT utilization had risen approximately 40% in the pilot regions and to the lesser extent of 18% in the control health regions (20). Although the authors seem to have most of the applicable data to isolate the proportion of the increase in screening which might plausibly be derived from their pilot, they did not attempt to do so. Primarily, data on gFOBT use from the pilot is based on a measure of utilization and does not reflect the adherence centered outcomes of interest to this project. There is also the issue of the experimental nature of the pilot program which may have influenced the practices of patients and physicians who were in a position to encourage screening in both registered participants and those who abstained from the pilot in the participating health regions. Although *ColonCancerCheck* utilizes many of the same interventions as practiced in this pilot, the differences in the actual implementation of these as they actually played out starting in 2008 is unclear and an investigation of *ColonCancerCheck* should use data from that period in which it was active.

Materials

A. Overview

The rationale of this thesis is to evaluate the effectiveness of a province-based organized colorectal cancer screening program at increasing the proportion of average risk individuals screened. As discussed in detail earlier, there is evidence indicating that CRC screening increased in several provinces including Ontario following the implementation of *ColonCancerCheck* (22, 23), which makes causal interpretation difficult. I make use of

a control group via cross-provincial comparisons in order to perform pre-post study designs which adjust for temporal trends in screening such as the Difference in Differences design, the Difference in Difference in Differences design, and an extension of the Regression Discontinuity Design. These causal models estimate the impact of *ColonCancerCheck* by comparing temporal contrasts in the outcome between the two treatment groups in order to control for temporal bias and time invariant differences between the treatment groups. The Canadian Community Health Survey is a useful tool in this context in that it is a nationally based survey and thus allows cross-provincial comparisons, it is largely consistent over time and permits use of repeated cross-sectional studies to measure values at multiple points prior to and following the intervention, and it contains survey questions on outcomes and covariates of interest to CRC screening in addition to other health and socio-demographic variables.

B. Canadian Community Health Survey

The data-source for this project is the confidential data-files from the Canadian Community Health Survey (CCHS) from the years 2003, 2005, 2007, 2008, and 2009 available from the Statistics Canada Research Data Centres. The CCHS is a nationally based survey which targets individuals aged 12 and older who are living in private dwellings in all Canadian provinces and territories. Persons who are not considered for the survey include those who live in Indian reserves, Crown lands, who are in institutions, are in the Canadian Forces, or live in very remote regions. The CCHS divides the provinces into health regions (HRs) and each territory is considered to encompass one health region. Allocation is organized to give a sufficient sample size for each HR with little disturbance to the proportionality of the allocation by province, and the sizes of the samples from each frame are increased before data collection to account for subject non-response, based on previous CCHS response rates (124).

The CCHS uses three sampling frame techniques to select households. An area frame derived from the Canadian Labour Force Survey frame was used to define strata and clusters and further define regions as major urban centers, cities, or rural regions. In CCHS 2009, 49% of the sample was collected through the area frame (124). A telephone list was used alongside nearly all the health regions to complement the area frame. In the

CCHS 2009, 50% of the sample was collected in the telephone frame. Finally, a random digit dialing frame was used to select households in four HRs in CCHS 2009 constituting 1% of that cycle's sample (124). The selection of individual respondents in the CCHS was designed to over-represent youths (aged 12-19). One person was selected per household using probabilistic measures to account for age and other types of household composition. Data was collected from January to December 2009 (for CCHS 2009) and computer assisted interviewing was implemented (124). Further information on survey design and implementation for the CCHS can be found elsewhere (124).

The result of the selection and weighting processes leads to the CCHS taking a complex design with stratification, multiple stages of selection, and unequal probabilities of selection for respondents. The difficulty this presented is that many weighting procedures, due to the sample survey framework, will provide correct estimates but incorrect variance calculations. Statistics Canada has made recommendations pertaining to variance calculation through using either coefficient of variation (CV) tables or bootstrap procedures. The bootstrap files are considered more robust as they can be used in a larger array of geographic levels which are not provided in CV tables, are able to calculate variance for parameter estimates from linear and logistic regressions, and are better suited to dealing with quantitative variables (124). Given the benefits of the bootstrap replication procedure and the availability of generated bootstrap replicates in appending files this method was chosen to adjust for variance calculation. All datasets were combined and modified and all calculations and analyses were conducted in STATA version 11 (125).

C. Survey Variance and Weighting

Surveys, such as the CCHS, with complex sample designs commonly have unequal probabilities of selection, variation in response rates across different subgroups, and other departures from the simple random selection model. Weights are often used to compensate for many of these features. Non-response classes are only used for those variables in which the researchers have an accurate estimate of everyone in the sample, including respondents and non-respondents. However, weighting for unit non-response is not capable of adjusting for biases derived from item non-response (126). Complex

samples often require more robust procedures to estimate sampling variance which take into account multistage sampling, weighting, and imputed values. Such techniques may include, in addition to the Bootstrap Replication technique already mentioned, the Taylor Series Approximation, Jackknife Replication, and Balanced Repeated Replication (126). Further details on variance composition in complex survey designs can be found in Appendix I part A: Survey Variance.

D. Bootstrap Re-sampling Technique

Here, I will briefly describe the underlying process and rationale for using bootstrap repeated replication techniques. In complex, multi-stage designs, sample reuse methods can be used as a means to approximate variance for nonlinear estimators $\hat{\theta}$ which may include ratios and coefficients. Reuse methods begin with a single sample from the population and estimate variance through a process of drawing repeated subsamples to create multiple estimates and calculation of the variance between these values. Reuse methods include balanced half samples, jackknife, and bootstrap techniques, and follow the general format:

1. K pseudo-samples are drawn from the data set
2. An estimate $\hat{\theta}_k$ mimicking the parent estimator $\hat{\theta}$ is drawn from each of the pseudosamples
3. V($\hat{\theta}$) of the estimator $\hat{\theta}$ is estimated by using the observed variation of $\hat{\theta}_k$ as in traditional variance estimates, based on $(\hat{\theta}_k - \hat{\theta})^2$. (127)

The bootstrap technique for variance estimation is different from the other methods mentioned primarily in the sampling of pseudo-samples from the original population. The method is to take a sample of clusters (≥2) from the strata of the original sample with replacement and to take a simple random sample of individuals from this group by strata. This is repeated *K* times to form *K* independent bootstrap samples (127).

The bootstrap resampling technique used in the CCHS has several steps in accordance to what has been discussed above. Replicates are selected and the variation between the

replicates is calculated. For each stratum in the sample, a simple random sample consisting of *(n-1)* clusters is selected with replacement to form a replicate which involves the recalculation of the weight of cluster with each replicate. Post stratification, using the same techniques which established the sampling design weights, is then performed to finalize the bootstrap weight replicates. This process is replicated for *B* times where *B*=500 (128). Further details on the formation of variance estimates and use of sample weights from bootstrap samples and evidence of the effectiveness of this technique can be found in Appendix I part B: Bootstrap Re-sampling Technique.

E. Combining Cycles of the Canadian Community Health Survey

Combining surveys is usually used in situations where there is either a need to increase sample sizes, examine an outcome effected by time trends in the population or improve coverage given in an individual survey (129). Given that this project is intent on examining temporal trends in proportions of average risk individuals screened according to province, we combined the five main cycles of the CCHS with available data on self reported colorectal cancer screening behaviours. No data from the sub-samples (sometimes labeled as the .2 cycles), which are smaller and focused on alternative health topics, were used. Various changes to the CCHS related to frequency and quantity of data collection, content, and survey methodology have been made over the seven year period of data collection representing periods before and after the implementation of *ColonCancerCheck* (130). The principal difference during this time was the redesign of the survey as a biennial survey with approximately 130 000 respondents (CCHS 2003, 2005 sometimes referred to as the .1 cycle surveys) to an annual survey with approximately 65 000 respondents (CCHS 2007, 2008, 2009). Changes have also been made with the annual survey design to the calculation of weights in order to adjust for this shift to annual survey cycles (130).

Covariate items were dropped from the analysis if they were not included in a sufficient proportion of the respondents to each survey year or if they presented in incompatible formats in different survey cycles. Unlike the CCHS 1.1 (2001) which is not used here, the area frame is used for approximately 50% of the sample in all surveys included so continuity based on this aspect of sample design was not felt to be an issue. This is of

significance as such variables as height, weight, physical activity and contact with physicians have been shown to be influenced by mode of data collection (131). Changes in the health regions across surveys are noted to be minor, and since this analysis is only interested at geography at the provincial level this is not anticipated to represent any notable difficulties (131).

Pooling of surveys is necessary to investigate the trends in screening behavior across time in Canada. Given consistency in the form and type of questions, and the consistency of the CCHS to represent the same population of interest across time, pooling is appropriate. In order to create estimates from a number of combined surveys, this project took the pooled approach as recommended by Thomas and Wannell for the CCHS (131). This approach pools cycles at the micro-data level to create a single large dataset. It is also necessary to combine and rescale the bootstrap weights in order to account for the combined cycles representing a single, manufactured, average population for correct variance estimation (131). Rescaling weights by a factor of j where j is the number of surveys pooled together will provide unbiased estimates of the total and has been recommended for use in the case of the CCHS (131) but has been cautioned against as it may result in inefficient estimations, particularly when the surveys pooled are quite different in size (129). Therefore, in order to rescale the original sample weights in an efficient manner, the weights of all five included CCHS surveys were adjusted by multiplying by a factor of $n_j/(n_1 + n_2 + \ldots n_j)$ where n represents the sample size of each survey $1,2\ldots j$. As each survey included relies on fundamentally the same framework of primary sampling units for data collection, it is not necessary to control for this potential difference in survey design (129).

Methods

A. Sample Characteristics

The study sample was designed to replicate, as closely as possible, the asymptomatic average risk population[3] which is the principal target of the mass screening program run by *ColonCancerCheck* (22) in addition to the similarly designed programs in development in Manitoba (121) and Alberta (122). As such, individuals were excluded from the dataset if they reported screening either with gFOBT or endoscopy for reasons related to family history of CRC, or due to having the procedure as follow-up to treatment of CRC. Individuals were also excluded if they reported having bowel disease such as colitis or crohn's disease. Age was restricted to 50-74 years in regards to the target group consistent within all three Canadian screening programs. The data was also divided into two geographical groups and two time periods of interest. The group organization created one group representing Ontario and one group representing individuals from all other Canadian provinces. I chose to not include information on reported screening practices from the three Canadian territories as these were seen as being poor counterfactuals of the province of Ontario and were for the most part poorly represented in the CCHS which resulted in very high coefficients of variation (17% to >33%) as reported previously by Wilkins and Shields (23). The two time periods created were indicators for the time before the implementation of *ColonCancerCheck* (2003, 2005, 2007) and after implementation of the program (2008, 2009).

Another important action was to restrict the data to individuals who had been surveyed the optional module of questions concerning screening for colorectal cancer. Due to a varying subset of provinces electing to include this module every year of the survey this created variation in the composition of the control group of provinces from year to year. *Table 1* indicates the participating provinces by year and Appendix II Part A indicates the proportional representation by province for each time period of interest.

[3] This may more accurately be described as the prior-to-first screening asymptomatic average risk population. This allows the sample to retain individuals who may be screening with endoscopy as a follow-up of a positive gFOBT who, as a result of this positive test, are not average risk.

Table 1: Participation in Module on CRC Screening Questions by Year

Province	2003	2005	2007	2008	2009
Newfoundland and Labrador	●	●	●	●	●
Prince Edward Island		●	●	●	●
Nova Scotia		●		●	●
New Brunswick		●		●	●
Quebec				●	
Ontario	○	●	●	●	●
Manitoba				●	
Saskatchewan	○		●	●	●
Alberta				●	
British Columbia	●			●	

○ *Indicates not all health regions were surveyed for this province. Since there was no evidence of disproportional representation of health regions which would influence the screening outcome measures the subsample of health regions used in the 2003 cycle was thought to be a suitable representative of the province.*

B. Descriptive Statistics and Outcome Modifiers

I first took mean comparisons of the entire covariate set to examine the differences in demographic characteristics, geography, healthy lifestyle behaviours, self-rated health measures and clinical features between Ontario and the remainder of provinces which answered the optional module on CRC screening in the CCHS. The purpose of this analysis was to examine the differences between Ontario and the other provinces in Canada which participated in the optional module survey for CRC screening. Subsequent analyses employed this group of provinces as a population-based counterfactual for Ontario and accounted for various differences between the two groups as an important step in justifying the exchangeability assumptions of the models. The results of these group comparisons can be found in Appendix II part B for the DD and DDD analysis. Appendix III Part A similarly shows treatment group differences for the regression discontinuity design. Additionally, period comparisons by province were made to examine how composition of dependent variables changed over time for the DD and DDD analysis and these results can be found in Appendix II Part C.

I then used univariate and multivariate logistic regression models to show the predictive capability of these covariates for the outcome variables denoting different types of

colorectal cancer screening in the past year through self-report. Other researchers (19, 23) have also used the CCHS in the past for the purposes of examining predictors of screening adherence in the Canadian context as has previously been discussed in Section II part F. Many of the covariates examined in this analysis were also used in these previous analyses. However, there are several important differences which are unique to this analysis. For one, rather than using a screening adherence measure which may include screening actions taken up to 5-10 years prior to answering the survey (in the case of colonoscopy) or two years for adherence to gFOBT, this analysis is primarily interested in past year screening only and is the only study to my knowledge which attempts to do so. Additionally this project uses more recent data from the CCHS 2009. A further addition of this project is to use data on prediction of screening as a means to draw comparisons between the provincial groups and between time periods.

C. Difference in Differences Design

In order to estimate the effect of *ColonCancerCheck* on the screening behavior of average risk individuals in Ontario aged 50-74, I employ a model most commonly used in the field of economics, the Difference in Differences (DD) design. The DD has sometimes been described as a version of a fixed effects model using aggregate data (132). This method should be appropriate to best measure the impact of a policy in a natural experiment which varies at a group (provincial) level in a non-randomized setting over time (132). The province of Ontario which receives the intervention is by default the treatment group and the other Canadian provinces, which are assumed to be sufficiently similar to the treatment group and received no similar intervention, are labeled as a single control group. As mentioned earlier, the Canadian territories are not represented in this sample. The DD design is structured as to remove bias due to the effect of temporal change throughout Canada in CRC screening practices and bias due to time invariant differences in screening between Ontario and other provinces. The framework of the population DD can be considered as follows:

	Intervention Group	Control Group
Pre-Intervention	$\beta 0 + \beta 1$	$\beta 0$
Post-Intervention	$\beta 0 + \beta 1 + \beta 2 + \beta 3$	$\beta 0 + \beta 2$

$$Y_{igt} = \beta 0 + \beta 1 * Group_g + \beta 2 * Time_t + \beta 3 * Group * Time_{gt} + \beta 4 * covar_{igt} + \varepsilon_{igt} \quad [1]$$

Where *Y* is the outcome of proportion of the average risk population screened for individuals *i* in province group *g* and time *t*. β3 is the effect measure, representing the screening outcome in the province of Ontario during and after 2008. Group or *G* is an indicator variable which indicates the province being Ontario, and Time or *T* is another indicator variable which denotes whether the cycle of the CCHS was 2008 or later. This corresponds with the formal launch of *ColonCancerCheck* in 2008 (22). Covariates are represented collectively by *covar* which includes variables for year, province, sex, age category (50-64 or 65-74), geography (rural or urban location), self-rated health (poor, fair, good, very good, excellent) reporting having a regular physician, reporting having had a flu shot, a categorized physical activity index score aggregating a number of leisure activities (active, moderately active, inactive), smoking status (regular, occasional, never), ethnicity (Caucasian or Other), education (>secondary school, secondary school, some post-secondary, post-secondary), income (household quintiles standardized to national level), and a categorized variable indicating the number of reported visits to a general practitioner in the past year (0-3, 4-10, 11-19, 20 or more). Of these, derived variables included household income (calculated from ratio of household income from all sources to community size and occupancy-specific low income cut-off for Statistics Canada and grouped into nationally based quintiles), and leisure-time physical activity (derived from questions pertaining to a number of forms of leisure physical exercise >15 minutes over the previous three months and reports of frequency of these activities). These covariates were selected in part due to their use in previous analyses of CRC screening using the CCHS (19, 23) and due to the results of *Table II* (See Appendix II Part D).Equation [2] represents the variable of interest.

$$\beta 3 = (\hat{E}[Y_{igt}|T=1, G=1] - \hat{E}[Y_{igt}|T=1, G=0]) - (\hat{E}[Y_{igt}|T=0, G=1] - \hat{E}[Y_{igt}|T=0, G=0]) \quad [2]$$

A key underlying assumption of this model is that, in absence of the intervention, the trends in the two treatment groups would be roughly similar (132). The intervention (*ColonCancerCheck*) is assumed to be the only factor which may cause a departure from

those trends and thus there is an assumption of no differential period effects at the time of intervention which may result in a change in screening behavior driven by an alternate cause. To the best of our knowledge, there has been no program introduction or budgetary actions initiated on a provincial level during this time period which would have any substantial impact on screening rates for CRC. The province of Manitoba during this time was undergoing a CRC screening program pilot from 2007-9 but this was limited to the regions of Winnipeg and Assiniboine (121) so this was not seen as something which could impact screening rates at a provincial level. Additionally, the province of Manitoba was only represented in the CCHS 2008 cycle which might reduce the impact of any marginal effects from this pilot. A screening program to be implemented in Alberta seems to have been in the planning phase during this time period with little evidence of concrete actions to increase screening activities (123).

A classic example using the DD design is the investigation of the effects of a change in minimum wage on employment in fast food restaurants. In April of 1992 the hourly minimum wage in the state of New Jersey rose by $US 0.80, at the time the highest in the country, and remained constant in the neighbouring state of Pennsylvania (133). Card and Krueger took data on employment in 410 fast food restaurants in the period before the intervention in February-March of 1992 and after the intervention in November and December of 1992. Fast food restaurants were chosen due to being a major employer of minimum wage workers, complying with minimum wage standards, generating a fairly homogeneous product given that only popular chains of restaurants were included. Employment in both states was assumed to be effected by the same trends in employment across that part of America. Contrary to conventional competitive models, the authors found that employment in New Jersey at the included restaurants did not decrease but rather had a small increase (133).

There are also examples of the Difference in Difference design used in the epidemiological and medical literature. Hollenbeak and colleagues (134) investigated the effect of intensive public reporting of hospital and physician performance on hospital mortality of six common medical conditions such as acute MI and Pneumonia. The authors used patients admitted in Pennsylvania hospitals, those affected by reporting

policies, and a propensity score matched cohort of patients from other states without public performance reporting environments. The time periods were prior to the reporting policies (1997-99) and after (2000-03). Results indicated lowered in-hospital mortality outcomes for Pennsylvania in 2000-03 with significant reductions seen in all outcomes and OR 95% CIs ranging from 0.59 [0.46, 0.76] for hemorrhagic stroke to 0.79 [0.67, 0.94] for sepsis (134). Another example of the use of this model in a medical context is Lairson and colleagues (135) who tested whether an intensified disease management program, largely based on patient monitoring, for patients enrolled in a large Houston-based clinic would have appreciable impacts for patients with Type-2 diabetes compared to the existing diabetes program in place at the clinic. Evaluating various outcomes relevant to diabetes care such as periodic testing adherence, glycemic control, costs and patient complications, the authors examined the differences which emerged between the group of patients who entered the more rigorous disease management program with the age, gender, an zip code matched controls who remained in the original program. Using this method, the authors found that most of the outcome measures were not statistically significant (135).

It is worth expanding on a few points in regard to the suitability of the Difference in Differences design to this thesis. In many cases here, the intervention does not have a single specific pathway by which it may affect the outcome. Individuals may decide to complete a screening test based on seeing advertisements, based on physician recommendations, based on increased knowledge of CRC, based on friends and relatives also screening, or some combination of these factors which *ColonCancerCheck* may influence. Therefore, there is no specific causal mechanism which is clearly defined by this process although the DD method isolates that effect due to the program as a whole. Card and Krueger's analysis of employment is another example of this lack of direct mechanistic link. There would be many reasons why a restaurant may change its hiring practices although the process infers that the resulting effect would be attributable to the minimum wage increase while controlling for all else (133). Also, it is worth pointing out that the presence of *ColonCancerCheck* does little to re-define the screening process or the standards of care as the same basic standards are relevant across Canada (6-8). Likewise, in both the case of the hospital public reporting policy (134) and the diabetes-2

disease management program (135) the outcomes affected by these interventions were not unique to the interventions but were desirable outcomes in regular clinical practice. The presence of the interventions in these cases is to help meet such standards, not to introduce new standards. The Lairson study is also an example of how two different groups may be compared using this methodology. In this study the cases and controls were distinguished by their healthcare plans which designated their ability to participate in one disease management program over the other (135). The methodological approach of controlling for group designation in the DD design allows for this notable difference to be accounted for which otherwise could have resulted in a severe selection bias. In the context of this study, despite the similarities between the provinces in terms of health care services delivery, it is clear from the literature (23) that individuals from Ontario historically have higher levels of screening compliance than the national average. The difference in differences design is able to control for this distinction as a time-invariant group level effect which would otherwise bias the results.

D. Difference in Difference in Differences Design

A subsequent analysis examined the demographic features of those screened to observe if there was a change in the profiles of those who received CRC screening in Ontario compared to those provinces without screening programs. Based on variables shown to be important modifiers of screening practices in average risk individuals aged 50-74 (see Appendix II Part D), I examined how the proportions of individuals screened in these demographic groups changed before and after program introduction in all provinces. I then completed cross provincial comparisons of the change in screening rates at the time of implementation of the screening programs, again using provinces without these programs as controls. An observed change in the rate of screening following the initiation of *ColonCancerCheck* in individuals of certain demographic groups in Ontario which differs significantly from the rate of change in control provinces would offer causal evidence that the program has made an outreach into specific groups. If such groups are less likely to receive screening otherwise, this would suggest *ColonCancerCheck* has been successful at increasing knowledge of and/or access to CRC screening. This model

could be described as the Difference in Difference in Differences (DDD) and follows a format similar to that above which can be summarized as follows:

$$Y_{igt} = \beta 0 + \beta 1 * Group_g + \beta 2 * Time_t + \beta 3 * Var_i + \beta 4 * Group * Time_{gt} + \beta 5 * Group * Var_{ig} + \beta 6 * Time * Var_{it} + \beta 7 * Group * Time * Var_{igt} + \beta 8 * covar_{igt} + \varepsilon_{igt} \quad [3]$$

Where *var* is the outcome modifying variable of interest and *β7* is the effect of interest, for example having a regular doctor, representing a change in the association of the modifying variable of interest in Ontario in the post intervention period.

This design has been used in the US to evaluate the impacts of a federal welfare policy on family composition in the context of state welfare policies implemented in different states and times (136). Another example of the use of this design exploits variation across states in the timing of the introduction of the Medicaid program and examines the impact on labour force participation on Medicaid-eligible single women throughout this time (137).

E. Regression Discontinuity Design

As the age of 50 is used to begin screening initiation into *ColonCancerCheck* (22) and the planned initiation age for future programs in Manitoba (121) and Alberta (123), this provides a potential threshold for a regression discontinuity design (RDD). Like the Difference in Differences model, RDD is a quasi-experimental design taking advantage of a naturally occurring treatment and control group. The RDD is applicable in instances where the probability of receiving treatment changes discontinuously as a function of an underlying variable, in this instance age. The discontinuity often exists due to some rule or policy pertaining to the intervention. This division also creates a comparable control group of individuals on one side of the intervention threshold who should be similar in most meaningful respects to the individuals in the treatment group on the other side of the threshold (132, 138).

When evaluating the success of a CRC screening program, this design is well-suited to capture any potential changes in screening rates in the average risk population which occur at the recommended age of 50 for screening. If screening recommendations are being followed, one would expect to see a significantly higher proportion of average risk

individuals at or just above the age of 50 being screened for CRC than those just below the age of 50 for which there is no program-specific targeting strategy.

However, in this case *ColonCancerCheck* may not be the only factor influencing screening beginning at age 50. National recommendations for screening of CRC are relevant in both Ontario and other provinces and consistently recommend screening for asymptomatic adults starting at age 50 (6-8). Previous research using a similar RDD design did not show significant threshold effects for most of the provinces including Ontario (96). I build on this earlier work by taking advantage of more recent information from the 2007- 2009 CCHS datasets which include colorectal cancer screening rates after the initiation of *ColonCancerCheck*. I also expand upon the basic RDD model by isolating the portion of the change in screening at the age threshold which may be attributed to *ColonCancerCheck* after adjusting for threshold effects over time and between groups.

The (fuzzy) RDD is focused on a treatment variable which is probabilistically conditional on an underlying variable, age. That is, there is a discontinuity in the probability of treatment at a certain age which is represented as a sharp change in the probability of screening at the age of 50 (x_0). This relationship between age and screening is not deterministic as in a sharp RDD, as the program does not mandate screening at and above age 50 and there are important elements of the program such as public advertising, which do not exclusively impact individuals over the age cutoff. It is useful to examine the outcome as an underlying conditional expectation function which is discontinuous at x_0.

$$E[Y_i|x_i]=E[Y_{0i}|x_i] + (E[Y_{1i}|x_i]- E[Y_{0i}|x_i])*Th_i \quad [4]$$

Where x_i is a continuous function of age centered at 0 at age 50, and Y is the outcome representing proportion screened. Th_i represents the treatment or threshold variable, being in the average risk group based on national recommendations and targeted by *ColonCancerCheck*, which is equal to 1 if $x_i \geq x_0$ and 0 if $x_i < x_0$. In equation 5, $E[Y_{0i}|x_i] = \beta 0 + \beta 01 * Age$ represents the expectation of the independent variable under conditions of no intervention, that is when $Th_i = 0$. This changes at the threshold ($x_i \geq x_0$) to create the new function $E[Y_{1i}|x_i] = Y_{0i} + \rho$ where ρ is the causal effect of interest, attributed to the

discontinuity in the treatment at the threshold. Substituting values into equation [4] and assuming that the relationship of screening with age may follow different polynomial distributions below and above the age threshold, we see the following equation:

$$Y_i = \beta 0 + \beta 01*Age_i + \beta 02*Age_i^2 + \ldots + \beta 0p*Age_i^p + \rho*Th_i + \beta 1*Age_i*Th_i + \beta 2*Age_i^2*Th_i + \ldots + \beta p*Age_i^p*Th_i + \beta 3*covar_{igt} + \varepsilon_i \quad [5]$$

This regression equation will be used to estimate the threshold effects attributable to screening guideline adherence generally for each of the four group-period combinations. There are several assumptions to this model (138). The first is that $P[Th_i|x_i]$ is discontinuous at x_0 so that there is a definite limit and $(Th_i = 1) \neq (Th_i = 0)$. It follows that the screening program or recommendations in question are the only reason for the discontinuity. Unlike a 'sharp' RDD, the assumptions that treatment is completely determined as a function of age do not apply although I use the Th_i variable in much the same way. In this case, I use this probabilistic increase in order to approximate a deterministic increase in treatment as a function of age. This is somewhat analogous to the intention to treat scenario whereby, although there is some degree of crossover to the treatment group, I perform the analysis as if individuals would adhere perfectly to their group allocation.

Secondly, in absence of the treatment, I assume that individuals close to the threshold are basically similar: $E[Y_{0i}| x_0 + \Delta > x_i \geq x_0] \approx E[Y_{0i}| x_0 - \Delta < x_i < x_0]$ where Δ is a small positive value. This particular assumption is more likely to hold if the analysis is restricted to individuals close to the threshold value, and is less likely as we include individuals farther away. I address this by using various bandwidths in the analyses and through covariate examination between treatment groups in Appendix III Part A. Indeed, evidence from previous studies on screening by age prior to *ColonCancerCheck* have indicated that older individuals are more likely to engage in screening (19, 23) but not that individuals within a few years of the cutoff are significantly different in this respect. A third assumption is that individuals cannot manipulate or select into treatment. Although individuals are ultimately responsible for seeking screening, the probability of treatment is a function of the screening program, which is not chosen by individuals, and obviously individuals cannot select their age.

In order to estimate the portion of effect which may be derived from *ColonCancerCheck*, I expanded on the RDD design to examine whether a differential threshold effect exists in Ontario after the intervention. By controlling for group-specific threshold effects and the interaction of these with time, this allows the model to control for any screening increases at the age cutoff which are plausibly derived from screening recommendations made earlier and isolates the marginal effect which I infer is a consequence of *ColonCancerCheck*. Model (6) follows a DDD format where the $\beta 7*Group*Time*Th_i$ term is the coefficient of interest representing the threshold effect in Ontario during the post intervention period, and can be written as follows:

$$Y_{igt} = \beta 0 + \beta 1*Group_g + \beta 2*Time_t + \beta 3*Th_i + \beta 4*Group*Time_{gt} + \beta 5*Group*Th_{ig} + \beta 6*Time*Th_{it} + \beta 7*Group*Time*Th_{igt} + \beta 8*Age_i + \beta 9*Age_i*Th_i + \beta 10*covar_{igt} + \varepsilon_{igt} \quad [6]$$

Apart from the example already discussed (96), the regression discontinuity design has been used frequently in the economics literature and to a lesser degree in the epidemiological literature. A simple example of the RDD is from Carpenter and Dobkin (139) who examine whether there is any change in fatal outcomes related to alcohol consumption in American young adults as they cross the threshold of 21 years of age, the legal drinking age in the US. In this example age in the continuous underlying variable, 21 is the threshold at which a policy is enacted, and the group of individuals aged 19-23 are felt to be representative of each other in most regards except for the inclusion of the policy. In this analysis, highly significant threshold effects of all cause morality, suicide, and motor vehicle accidents were seen and attributed to threshold effect on legal drinking age (139).

F. Placebo Tests: Checks of Test Suitability and Bias

I also employ a number of verification tests of the previous procedures to check the suitability of the tests and to investigate bias. In checking the DD methods, I perform a DD within the period before intervention with the year 2007 being labeled as a post-intervention time period. A positive result of the effect measure in this analysis may indicate the results are biased, possibly by uncontrolled trends in screening which are not accounted for in the DD model. I also perform the DD on an outcome which is not meant

to be affected by *ColonCancerCheck*: reporting having taken a flu shot in one's lifetime; again, if the result of this test is significant then the DD may be biased. Similarly, the DDD test will be tested by utilizing an interaction which should not be effected by the screening program: in this case those whose age is below the screening threshold for those of average risk. I tested the RDD for bias by setting the threshold age to different levels, to investigate the robustness of the model and whether distortions in screening trends by age may be strong enough to create a significant effect. These methods of bias-testing, sometimes known as 'placebo tests' where the test examines whether there is an effect where there is no reason for there to be one, are derived from recommendations made by Duflo (140).

Results

A. Comparison of the Treatment and Control Groups for DD Effect Estimation

Restricting the sample to those aged 50-74 in Canadian provinces participating in the optional module of CRC screening, the initial size of the sample was 81 262, and then was reduced to 71 232 after excluding individuals reporting screening due to family history or as part of treatment or having colitis, and reduced to 58 142 after restricting the data for complete case analysis. Of this, Ontario contributed 19 888 in the pre- and 11 293 in the post-intervention period and the remainder of Canada contributed to 12 324 in the pre- and 14 637 in the post-intervention period. Results suggest that the effects of item non-response were not informative of the outcome. Analyses used in Part A and Part B of this section use a more restricted sample based on complete case analysis with a larger covariate set with a national sample size of 55 444.

Appendix II Part A shows the weighted proportion of asymptomatic average risk individuals aged 50-74 in each time period and each Canadian province. Ontario, the intervention group, accounts for approximately 71% of the study population in the pre-intervention period and 51% of the population in the post-intervention period. Notably, there is no representation from the provinces of Quebec, Manitoba or Alberta in the pre-

intervention time period. The Atlantic Provinces (Newfoundland and Labrador, Prince Edward Island, Nova Scotia, New Brunswick) together constitute approximately 12% of the sample in the pre-intervention period and 10% of the sample in the post intervention period. The province of Quebec is only represented in the CCHS 2008 although still constitutes the second highest proportion of the sample in the post intervention period at close to 18%. British Columbia constitutes nearly 14% of the sample in the pre-intervention period, making it the second most represented province behind Ontario, although only 9% in the post intervention period.

In Appendix II Part B the tables show comparisons of variables between the intervention and control groups. There is a consistent difference between the two groups in regards to geographical location where approximately 83% of respondents from Ontario and around 70% of the control population report living in an urban area. Neither the age or sex compositions of the two groups are significantly different from one another. In both time periods it is evident that Ontario has a substantially lower proportion of people born in Canada at just above 60% than the control provinces which have above 80% of their respondents Canadian born. Differences are most notable in the Canadians who have immigrated from South and Central America and the Caribbean, Europe, and Asia. Although the proportions of recent immigrants (0-9 years since immigration) to the two groups are similar, Ontario seems to have nearly three times as many long-established immigrants (20 or more years in Canada) than the control group in the post intervention but not pre intervention period. Linking these trends to ethnicity it is apparent that Ontario consistently shows a higher proportion of visible minorities (non-Caucasians).

In terms of education and household income Ontario seems to skew towards having larger proportions in the highest categories in the pre-intervention period although this difference seems to diminish in the later period. There are no significant differences in terms of marital status in the pre-period although the post-intervention period does show differences with higher proportions of Ontarians reporting being in marriages. Examining a number of health and lifestyle variables, there seems to be no consistent or strong significant differences between the groups at either time period in regards to body mass

index categories, smoking status, physical activity, frequency of heavy alcohol consumption, self-perceived general health, mental health or stress.

In the pre-intervention period there is a significant but probably not substantial difference in the number of individuals who report having a regular doctor. In the post-intervention period this difference seems to have increased to nearly 7% point difference with close to 95% of respondents from Ontario and 88% of respondents from other provinces reporting having a regular doctor. Further examination of reported physician visits show no very strong patterns in favour of higher numbers of general practitioner (PCP) or specialist consultations in the past year by group. There may be some weak trends in favour of greater numbers of PCP visits for the control group in the pre-intervention period which are somewhat reversed in the latter period, and a weak trend towards greater number of reported specialist visits in respondents from Ontario. One of the starkest differences in regards to medical services use is that there is a much higher proportion of respondents from Ontario reporting having had a flu shot than the control provinces which may relate to a previous public health policy. Strong and significant differences between the groups were also observed in the outcome variables including past year screening and the more common past 2-5 year screening measures. These results further reinforce that screening in Ontario is higher for both the gFOBT tests and that the proportion screened in both groups is increasing. Appendix II Part C further shows that this shift in all reported screening outcomes is positive and significant in both groups for all measures of the screening outcome. However, these results are still insufficient to establish that the presence of an organized colorectal cancer screening program in Ontario has had an impact.

The difference in differences design allows for the control of fixed differences between the province of Ontario and the other Canadian provinces serving as controls. Additionally, I include several of these variables as covariates in the DD model to further control for these between-group differences. Providing these do not change in a pattern parallel to that of *ColonCancerCheck* this model should account for most of the differences between the two groups.

B. Predictors of Screening

I next examined the predictors of past-year screening for colorectal cancer. Appendix II Part D shows the marginal effects for each covariate as calculated from data from the four group-period pairings. The dependent variable was set as a self-report of having a gFOBT or endoscopy in the previous year. Consistent predictors of screening across both the intervention and control groups in this population in the pre-intervention period included being in the higher age category of age 65-74, being non-smoker, being physically active, reporting having had a flu shot, reporting having a regular medical doctor and visiting the doctor more often. In the pre-intervention period, in particular, the effect of having a regular medical doctor was very predictive with a marginal effect of 0.10 95% CI [0.044, 0.158] in Ontario and 0.069 [0.024, 0.113] in the control provinces. Reporting having taken a flu shot was also a strong predictor of screening at 0.065 [0.043, 0.087] for Ontario and 0.033 [0.016, 0.049] in control provinces. Advanced age of 65-74 was a slightly less strong predictor of screening at 0.027 [0.007, 0.047] in Ontario and 0.049 [0.029, 0.069] in control provinces.

In the pre-intervention period, outside of Ontario only, male sex was a significant but weak predictor of screening at 0.025 95% CI [0.008, 0.041] and identifying as Caucasian is an even more strongly associated at 0.049 [0.026, 0.073]. Additionally, those with higher levels of obesity (class II and III) as measured by BMI self-report are significantly less likely to screen and those who self-rate their overall health as excellent are more likely to screen. In Ontario only, those who report being separated (-0.070 [-0.109, -0.031]) or are single individuals who never married (-0.034 [-0.063, -0.006]) were less likely to screen than married individuals. There were also significant increases seen in screening for those with some post secondary (0.062 [0.016, 0.107]) and post secondary (0.028 [0.005, 0.050]) levels of education compared to less than secondary school education and those who report finding stress levels as extremely stressful are also more likely to screen (0.081 [0.016, 0.146]). Not shown are associations on country of birth and screening with Canada being the reference standard. Data was only available in sufficient quantity for Ontario but not for the control provinces. There were no significant associations between country of birth and screening behavior.

I also examined predictors of screening in both groups in the post intervention period. For the most part these did not differ substantially from the associations seen in the pre-intervention period. In Ontario an increase in screening likelihood compared to the pre-intervention period was observed for being an older age 65-74 (0.060 [0.030, 0.090]), having had a flu shot (0.095 [0.061, 0.128]), in addition to an increase in association between screening and having a regular medical doctor (0.282 [0.214, 0.350]) by nearly threefold. Education level, smoking status, marital status, and stress were not associated with past year screening in Ontario in the post intervention period although in most cases these changes don't represent a large change from the pre-intervention associations. In the control provinces there seemed to be less change in the strength of the associations over time. Unlike Ontario associations of the screening outcome and age did not change by very much (0.042 [0.020, 0.063]) and the associations of sex and ethnicity to screening also remained fairly constant. Being a non-smoker and rating one's health as excellent also ceased to predict screening behaviour and the association between having a flu shot and screening increased only slightly (0.047 [0.028, 0.066]). Notably, the association between screening and having a regular medical doctor nearly doubled (0.125 [0.075, 0.175] from the earlier time period.

I also reproduced all of the multivariate results with univariate results as a means to test for confounding (not shown). In almost all cases the values in the univariate and multivariate analysis are quite similar. The main difference between these models seems to be that the importance of having a regular doctor is somewhat exaggerated in the univariate model and thus might be somewhat confounded by other attributes. However, even in the multivariate model this variable is a very strong and consistent predictor of screening even after accounting for many of the other variables it may be associated with.

C. Difference in Differences Design Outcomes

Table 2 shows the results of the difference in differences regressions. Indicated are the results of the indicator terms for the post intervention time period, the intervention group (Ontario), and the interaction term consisting of these two which denotes the marginal effect of *ColonCancerCheck* on screening in the average risk population.

Table 2: Difference in Differences Design Output

Outcome		DD*				DD**			
		Mrg Effect	95% CI		p	Mrg Effect	95% CI		p
Combined	**Post Intervention**	0.090	0.066	0.114	0.000	0.090	0.067	0.114	0.000
	Intervention Group	0.087	0.064	0.111	0.000	0.063	0.039	0.086	0.000
	Group*Time	0.052	0.029	0.074	0.000	0.055	0.033	0.077	0.000
gFOBT	**Post Intervention**	0.073	0.052	0.095	0.000	0.074	0.053	0.095	0.000
	Intervention Group	0.082	0.059	0.104	0.000	0.064	0.041	0.086	0.000
	Group*Time	0.050	0.030	0.070	0.000	0.052	0.032	0.071	0.000
Endoscopy	**Post Intervention**	0.022	0.008	0.037	0.003	0.023	0.009	0.038	0.002
	Intervention Group	0.018	0.005	0.031	0.006	0.009	-0.004	0.022	0.182
	Group*Time	0.007	-0.007	0.021	0.301	0.009	-0.005	0.023	0.207

Using CCHS 2003, 2005, 2007, 2008, 2009
N=58142, Weighted N= 2882630. All values are weighted.
Dependent variable is self-report of having a gFOBT in past year or endoscopy (flexible sigmoidoscopy or colonoscopy) in past year or combined measure
**: controlled for year and province indicators*
***: controlled for year, province, sex, age category, geography, self-rated health, having MD, reporting flu shot, physical activity index, smoking status, ethnicity, education, income, #GP consultations past year*
Results shown are average marginal effects calculated from multivariate logistic regression model
Marginal Effects are based on valid responses (complete case analysis used)
Marginal Effects based on closest approximation of average risk population available ages 50-74

A partially controlled model and a more fully controlled model were implemented for each of the three outcomes. The complete results for the gFOBT and endoscopy outcomes under the fully controlled DD model can be found in Appendix II part E. The marginal effect in the fully controlled model of the time indicator on gFOBT screening is 0.074 95% CI [0.053, 0.095] and 0.023 [0.009, 0.038] on screening by endoscopy. There is also a significant positive effect observed on the influence of Ontario on screening for gFOBT outcomes (0.064 [0.041, 0.086]) although there is no evidence on screening with endoscopy (0.009 [-0.004, 0.022]). Since the marginal effect of this variable decreases from 1.8 to 0.9 percentage points upon introducing the covariate set confounding of this variable is probable. Represented by the *Group*Time* interaction term, estimated increases in the proportion of asymptomatic average risk individuals who screened with gFOBT in the past year due to *ColonCancerCheck* are 0.052 95% CI [0.032, 0.071] or a 5.2 absolute percentage point increase in proportion screened. I also find no effect of the program on increasing endoscopy outcomes 0.009 [-0.005, 0.023].

Figure 1: Past Year Screening in Canada over Time Period

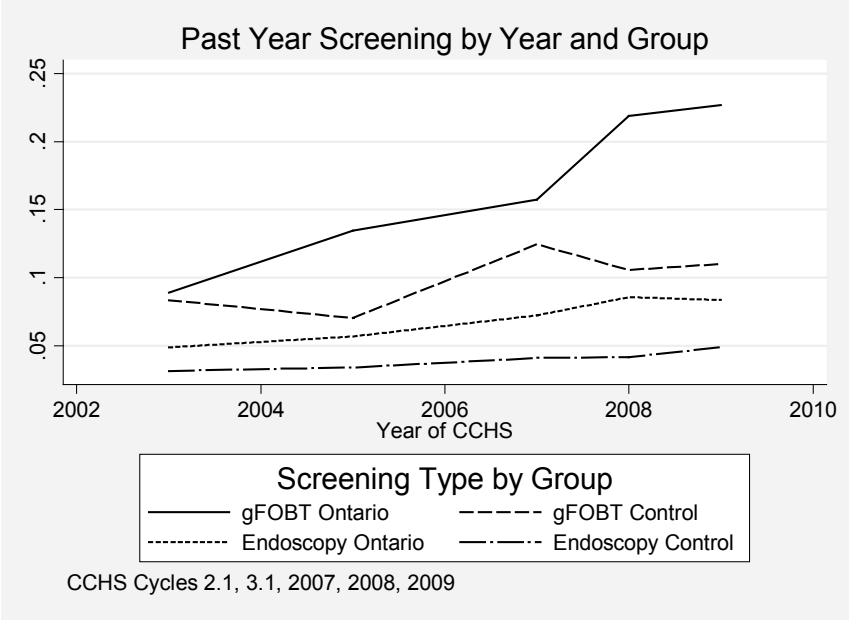


Figure 1 illustrates past year screening over time for each of the gFOBT and endoscopy outcomes for both the intervention and control groups. It is evident that there is an increasing trend in both outcomes, that there are greater numbers of adherents of gFOBT than endoscopy, and that for both screening outcomes Ontario has greater levels of adherence than elsewhere in Canada. gFOBT screening in Ontario appears to increase gradually until 2007, at 15.74% adherence, where it sharply increases to 21.90% at the year 2008, the year of the implementation of *ColonCancerCheck,* and continues at a more gradual rate of increase to 22.67% in 2009. gFOBT screening in the control provinces varies somewhat from year to year, likely result of the alterations in provinces which constitute the control group. For instance, in 2005 when the gFOBT screening adherence decreases to 7.03% from 8.34% in 2003, the control group is comprised of only the four Atlantic Provinces, which tend to have screening rates below the national average (23).

D. Difference in Difference in Differences Design Outcomes

The importance of several covariates which predicted the outcome was seen to change over time as discussed in Part B of this section. Some of these modifications in the strengths of association between these covariates and the outcome are possibly attributable to *ColonCancerCheck*. For example, some factors that are integral to

disseminating knowledge about CRC screening in the program (having access to a regular medical doctor) seem to have increased in importance. The strengths of the associations observed in regards to advanced age, having a flu shot, having a PCP, and physical activity to the combined screening outcome were all observed to change over time either only in Ontario (age, physical activity) or disproportionably so (flu shot, PCP). It was also possible that *ColonCancerCheck* may have granted access to screening technologies to individuals who otherwise would not have engaged in screening. Below in *Table 3* are the results of Difference in Difference in Differences analysis (see Equation [3] where β7 is the effect measure) indicating the effect of interest by interacting the group and time variables with the modifying covariate.

Table 3: Difference in Difference in Differences Output

Outcome		**DDD***			
		Mrg Effect	**95% CI**		**p**
gFOBT	**G*T*MD**	0.008	-0.095	0.111	0.883
	G*T*Age65-74	0.028	-0.008	0.065	0.127
	G*T*PhysInactive	0.007	-0.029	0.042	0.720
	G*T*Flu-Shot	0.010	-0.031	0.052	0.623
Endoscopy	**G*T*MD**	0.065	-0.007	0.137	0.077
	G*T*Age65-74	0.002	-0.022	0.027	0.853
	G*T*PhysInactive	-0.003	-0.027	0.022	0.835
	G*T*Flu-Shot	-0.007	-0.035	0.021	0.628

Using CCHS 2003, 2005, 2007, 2008, 2009
N=58142, Weighted N= 2882630. All results are weighted.
Dependent variables are self-report of having a gFOBT in past year or endoscopy (flexible sigmoidoscopy or colonoscopy) in past year. G is a group indicator=1 if province is Ontario and 0 otherwise, T is a time indicator =1 if year≥2008 and 0 otherwise. Modifying variables include MD for reporting having a regular medical doctor, Age65-74 for being aged 65 to 74 years, PhysInactive is an indicator for those who self-report engaging in physical activity very infrequently, and Flu-Shot is an indicator for those reporting having had a flu shot.
**: controlled for year, province ,Group, Time, Group*Time, Group*Var, Time*Var, sex, age category, geography, self-rated health, having MD, reporting flu shot, physical activity index, smoking status, ethnicity, education, income, #GP consultations past year. Note: var is the modifying variable of interest*
Results shown are average marginal effects calculated from multivariate logistic regression model
Marginal Effects are based on valid responses (complete case analysis used)
Marginal Effects based on closest approximation of average risk population available ages 50-74

There is no significant evidence for a change in the importance of the associations between these variables and screening upon the introduction of *ColonCancerCheck* at the 95% confidence level. However, two of the estimated effects are large enough in

magnitude that they warrant further discussion. I estimate a nearly three percent absolute increase in the strength of association of being aged 65-74, although this association is not strong enough to indicate significance at this level. Secondly, although *ColonCancerCheck* was not observed to increase past year screening by endoscopy in this time frame, there does seem to be a considerable increase in the importance of having a regular medical doctor associated with past year screening endoscopy of over 6%. This increase in the strength of association, although nonsignificant, does hint at the possibility of a change in the mode of delivery of these forms of services which occurred alongside *ColonCancerCheck*.

Part of the outputs for the DDD tests including the MD interaction can be found in Appendix II Part F Although it is evident that having a regular physician is very important for the process of obtaining a gFOBT, and that the strength of this association on a linear scale was observed to increase in Ontario over time, the results indicate fairly strongly that there was no meaningful change in this association. It is probable for many of these findings that there was insufficient sample size available to be certain of the strengths of these interactions.

Having some level of post-secondary education and being a non-smoker were each associated with screening in the pre-intervention period in Ontario, and then ceased to be associated with the combined screening outcome in the post intervention period. However, being a non-smoker followed the exact same pattern in the control provinces suggesting that this was a national trend rather an impact of *ColonCancerCheck*. Additionally, higher levels of education were not associated with screening in either the pre or post intervention periods outside of Ontario, which creates difficulty inferring program influence in the case that the association differed pre-intervention and then became more like that of other provinces after intervention. Given these associations, I decided not to further investigate the education and smoking covariates.

E. Regression Discontinuity Design Effect Estimation

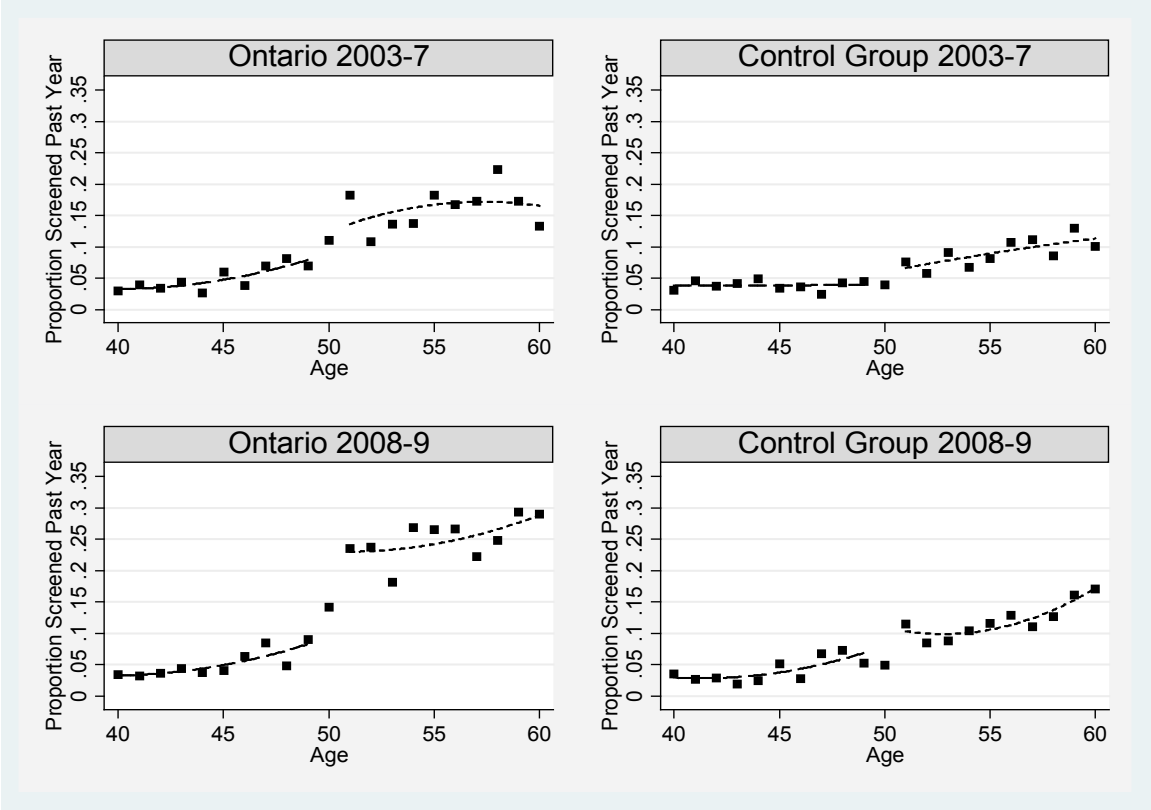
Appendix III Part A shows the distributions of baseline covariates on either side of the age 50 threshold. I compare two distributions of different bandwidths, that is, different

ranges of participant age. Since the individuals below the age threshold are meant to serve as controls for those above it, examining various characteristics of the sample are a way to analyze the suitability of this group as a counterfactual. In most cases the divisions formed through treatment group allocation create fairly balanced groups, with many of the differences relatively minor and arguably unsubstantial. Differences that do exist are consistently reduced with the smaller bandwidth. The treatment group is seen to consistently have a greater number of married individuals, fewer normal weight individuals, greater numbers of non-smokers, greater numbers of inactive individuals and somewhat fewer individuals rating their overall health as excellent although none of these differences represent a very large disparity. Very similar proportions of individuals report having a regular medical doctor at approximately 89% below the threshold and 92% above it. A slight pattern emerges where the treatment group tends to report higher numbers of physician visits. The most apparent increase is in the proportion of individuals reporting having taken a flu shot in the above-cutoff group.

A concern in this design is that expanding the bandwidth, while increasing precision, will exacerbate the differences between the groups divided by the cutoff and increase the likelihood of other uncontrolled factors influencing the dependent variable. In this sample it appears that the differences are mostly minute and that there is a strong likelihood that the below cutoff group can serve as a counterfactual for the group above the cutoff conditional on selected covariates. I also examined the distribution of various dependent variables in this context and there are strong increases in reported screening at the age cutoff for both types of screening.

Figure 2 shows screening by age in the range of 40-60 years with screening defined as reporting having a gFOBT or endoscopy in the previous year. The population is the approximation of the asymptomatic average risk population, all proportions are weighted, and there are two quadratic fit terms for each graph based on participants being above or below the cutoff age of 50. An examination of these graphs displays quite vividly that the patterns of past year screening around the cutoff change noticeably over time in each group and that this change is most noticeable in Ontario.

Figure 2: Age and Past Year Screening by Group and Time Period



In the pre-intervention period in the control provinces there is almost no discernible jump in screening at the threshold although the slope definitely appears to increase. In the post-intervention period in the control provinces there is more evidence of screening participation increasing directly below the cutoff and there being a small jump at the cutoff. This pattern is also an apt description of Ontario in the pre-intervention period although screening levels are higher in general. Ontario in the post intervention period seems to show the highest discontinuity jump at the threshold age indicating an increase from 8.98% participation at age 49 to a 23.6% participation at those of age 51. Appendix III Part B indicates the precise levels of screening at each age which are represented in *Figure 2*.

Notably, in all time and group pairings the largest jump in reporting the combined past year screening outcome is in individuals aged 51. In the interest of formulating the correct polynomial term for age in the regression discontinuities, I ran two basic predictive models using a linear and squared age term to analyze how well these models were able

to capture changes in the outcome at values at close distance to the cutoff as recommended by Angrist and Pischke (132). Both models appear to be fairly good fits, and simply examining *Figure 2* it would seem that within this age range a linear relationship between age and the outcome with the presence of a discontinuity is a very close approximation to the fit offered by a higher order term.

Table 4 indicates models both utilizing linear and squared terms to represent the continuous relationship of age with and without a set of controls. Comparisons of results using the limited control set (year and province) and the full set do not show very large departures indicating that the age cutoff in the model was largely able to mimic a randomization procedure. When using a model with either a linear or squared age trend there were significant results observed at 95% confidence for screening at the treatment threshold in Ontario in the post-intervention period with gFOBT as the dependent variable. These results were higher in Models 1 and 2 and represented an absolute increase in screening at the threshold of 9.10 95% CI [4.33, 13.87] and 8.42 [3.81, 13.03] percentage points, respectively. In models 3 and 4 these results diminished somewhat to 7.65 [0.53, 14.76] and 7.12 [0.10, 14.14] respectively.

In Ontario in the pre intervention period there was an observed increase in screening by gFOBT between an absolute increase of 1.01 to 2.64 percentage points, depending on which model was used, and none of these increases were large enough to be significant at a 95% level. In both periods, the control group did not exhibit a significant threshold effect for gFOBT and in some cases there was some suggestion of a negative effect in screening of this type at the threshold by a small degree. Only very minor and non-significant threshold effects were observed for endoscopy in all group and time combinations.

Table 4: Regression Discontinuity Design

			Model 1				Model 2			
			Mrg Effect	95% CI		p	Mrg Effect	95% CI		p
gFOBT	**Ontario**	**2003-7**	0.0264	-0.0016	0.0544	0.064	0.0229	-0.0045	0.0503	0.101
		2008-9	0.0910	0.0433	0.1387	0.000	0.0842	0.0381	0.1303	0.000
	Control	**2003-7**	0.0072	-0.0162	0.0306	0.548	0.0059	-0.0172	0.0291	0.617
		2008-9	-0.0143	-0.0451	0.0165	0.364	-0.0118	-0.0414	0.0178	0.435
Endoscopy	**Ontario**	**2003-7**	0.0173	-0.0026	0.0371	0.089	0.0162	-0.0030	0.0355	0.098
		2008-9	0.0044	-0.0280	0.0368	0.792	0.0058	-0.0260	0.0376	0.722
	Control	**2003-7**	0.0090	-0.0092	0.0272	0.332	0.0085	-0.0091	0.0262	0.344
		2008-9	0.0162	-0.0038	0.0362	0.113	0.0129	-0.0067	0.0325	0.198

			Model 3				Model 4			
			Mrg Effect	95% CI		p	Mrg Effect	95% CI		p
gFOBT	**Ontario**	**2003-7**	0.0139	-0.0310	0.0589	0.543	0.0101	-0.0341	0.0543	0.654
		2008-9	0.0765	0.0053	0.1476	0.035	0.0712	0.0010	0.1414	0.047
	Control	**2003-7**	-0.0013	-0.0440	0.0414	0.952	-0.0037	-0.0450	0.0376	0.861
		2008-9	-0.0261	-0.0708	0.0186	0.253	-0.0243	-0.0676	0.0190	0.272
Endoscopy	**Ontario**	**2003-7**	-0.0025	-0.0336	0.0286	0.874	-0.0034	-0.0338	0.0270	0.825
		2008-9	0.0001	-0.0580	0.0582	0.997	0.0020	-0.0545	0.0585	0.944
	Control	**2003-7**	0.0012	-0.0328	0.0351	0.946	0.0021	-0.0299	0.0342	0.896
		2008-9	0.0048	-0.0211	0.0307	0.716	0.0030	-0.0225	0.0286	0.816

Using CCHS 2003, 2005, 2007, 2008, 2009
Ontario 2003-7: N=19813,Weighted N=1508273.7;
Ontario 2008-9: N=10334,Weighted N=847378.05;
Control 2003-7: N=12419,Weighted N=5591854.64,
Control 2008-9: N=13489,Weighted N=780309.78
All results are weighted. Bandwidth is ages 40-60.
Model 1: adjusted for age, ageTh_i, year, province*
Model 2: adjusted for age, ageTh_i, year, province, sex, geography, self-rated health having regular MD, reporting flu shot, physical activity index, smoking status, ethnicity, education, income, #GP consultations past year*
Model 3: adjusted for age, ageTh_i, age^2, age^2*Th_i, year, province*
Model 4: adjusted for age, ageTh_i, age^2, age^2*Th_i, year, province, sex, geography, self-rated health having regular MD, reporting flu shot, physical activity index, smoking status, ethnicity, education, income, #GP consultations past year*
Dependent variable is self-report of having a gFOBT in past year or endoscopy (flexible sigmoidoscopy or colonoscopy) in past year. Marginal Effects represented are for treatment variable (Th_i)
Results shown are average marginal effects calculated from multivariate logistic regression model
Marginal Effects are based on valid responses (complete case analysis used)
Marginal Effects are based on closest approximation of average risk population available

Several additional models of these treatment effects were analyzed and are not shown here. It was decided that the additional restrictions brought on by the limited bandwidth of 45-55 years of age were unnecessary given the suitability of the 40-60 range in approximating randomized groups and the deficiency in precision brought about by the

narrow age range. Similarly, employing third degree polynomial terms for age greatly diminished precision and did not seem warranted given the near linear relationship observed in *Figure 2* and the evidence for model fit from Appendix III Part B.

F. *ColonCancerCheck* Threshold Effect Estimation

As discussed in Methods Part E, the findings of a significant threshold effect as was observed in the post-intervention period in Ontario may be attributed either to *ColonCancerCheck*, which recommends screening the asymptomatic average risk population at the age of 50, or through compliance with a screening guidelines generally which would impact a certain proportion of the population in absence of the program. The following model, formally presented in Equation [6], controls for threshold effects similarly to the DD model to control for time invariant group-specific effects and time trends in regards to screening at the age threshold of 50.

The results of these DDD tests indicate that there is insufficient evidence to make the claim that *ColonCancerCheck* has increased past year screening at the threshold for any of the model types at a 95% confidence level. It is worth pointing out, however, the size and near significance of the marginal effects calculated for the gFOBT outcome in both the Models 3 and 4 for which the bandwidth is 45-55 years of age. Models 3 and 4 indicate a 4.56 95% CI [-0.40, 9.38] and 4.51 [-0.35, 9.38] percentage point increase in past year gFOBT screening at the threshold respectively, which is, although not conclusive, very suggestive of a considerable treatment effect. Models which use the less restrictive bandwidth of 40-60 years of age show a substantially weaker and also non-significant treatment effect of 1.04 [-2.45, 4.53] and 1.19 [-2.25, 4.63] for Models 1 and 2 respectively.

Table 5: Threshold Effect of ColonCancerCheck Estimations

Outcome	Model	Mrg Effect	95% CI		p
gFOBT	1	0.0104	-0.0245	0.0453	0.558
	2	0.0119	-0.0225	0.0463	0.499
	3	0.0456	-0.0040	0.0953	0.072
	4	0.0451	-0.0035	0.0938	0.069
Endoscopy	1	-0.0067	-0.0283	0.0149	0.544
	2	-0.0063	-0.0278	0.0152	0.566
	3	-0.0103	-0.0396	0.0189	0.489
	4	-0.0097	-0.0386	0.0192	0.510

Using CCHS 2003, 2005, 2007, 2008, 2009
N=56055, Weighted N= 3727816.1 (40-60)All results are weighted.
N=28541, Weighted N= 1969611.2 (45-55) All results are weighted.
*Model 1: Controlled for Group, Time, Th_i, Group*Time, Group*Th_i, Time*Th_i, year, province, age, age*Th_i Bandwidth 40-60.*
*Model 2: Controlled for Group, Time, Th_i, Group*Time, Group*Th_i, Time*Th_i,year, province, sex, geography, self-rated health having regular MD, reporting flu shot, physical activity index, smoking status, ethnicity, education, income, #GP consultations past year, age, Age*Th_i Bandwidth 40-60*
*Model 3: Controlled for Group, Time, Th_i, Group*Time, Group*Th_i, Time*Th_i, year, province, age, age*Th_i Bandwidth 45-55*
*Model 4: Controlled for Group, Time, Th_i, Group*Time, Group*Th_i, Time*Th_i, year, province, sex, geography, self-rated health having regular MD, reporting flu shot, physical activity index, smoking status, ethnicity, education, income, #GP consultations past year, age, age*Th_i Bandwidth 45-55*
*Dependent variable is self-report of having a gFOBT in past year or endoscopy (flexible sigmoidoscopy or colonoscopy) in past year. Marginal Effects represented are for $\beta 7$*Group*Time*Th_i*
Th_i represents the treatment variable and is equal to 1 if age≥50 and 0 otherwise
Results shown are average marginal effects calculated from multivariate logistic regression model
Marginal Effects are based on valid responses (complete case analysis used)
Marginal Effects based on closest approximation of average risk population available

Results for endoscopy are consistently non-significant between 0 and -1.0 percentage point differences where all results are consistently negative. The strongest results are from Model 3 and point towards a 1.03 [-3.96, 1.89] percentage point decrease in past year endoscopy. Similar models to those above were also carried out which included a squared age polynomial term with the interactions of age with the cutoff term (Th_i). These are not shown here but provided nearly identical results to those above.

G. Placebo Tests

Several additional tests on the same populations used in the difference in differences and regression discontinuity designs were implemented to test for bias, which would be indicated by significant results for an alternate and unrelated dependent variable or a time specification which did not represent any intervention period or relevant threshold age.

Table 6: Difference in Differences Bias Check

Part A

Outcome		DD*				DD**			
		Mrg Effect	**95% CI**		**p**	**Mrg Effect**	**95% CI**		**p**
gFOBT	Post Intervention†	0.0865	0.0533	0.1196	0.000	0.0875	0.0551	0.1200	0.000
	Intervention Group	0.0844	0.0600	0.1088	0.000	0.0688	0.0439	0.0936	0.000
	Group*Time	-0.0191	-0.0549	0.0167	0.295	-0.0204	-0.0557	0.0148	0.256
Endoscopy	Post Intervention†	-0.0006	-0.0216	0.0205	0.959	0.0001	-0.0205	0.0208	0.989
	Intervention Group	0.0186	0.0036	0.0336	0.015	0.0100	-0.0052	0.0253	0.198
	Group*Time	0.0220	-0.0014	0.0454	0.065	0.0218	-0.0010	0.0447	0.061

Part B

Outcome		DD*				DD**			
		Mrg Effect	**95% CI**		**p**	**Mrg Effect**	**95% CI**		**p**
Flu Shot	Post Intervention	0.0661	0.0364	0.0958	0.000	0.0743	0.0453	0.1032	0.000
	Treatment Group	0.2725	0.2502	0.2949	0.000	0.2588	0.2366	0.2810	0.000
	Group*Time	-0.0153	-0.0445	0.0139	0.305	-0.0153	-0.0436	0.0131	0.291

Model Specification is the same as in Table 2 unless stated otherwise
In Part A: N=30484, Weighted N= 1468618.9
†: The post intervention period has been altered to equal 2007. Data from 2008-9 are removed.

For the placebo tests of the DD models, I used a part of or the full equivalent of the population examined in *Table 2* including placing the same restrictions on the sample and including the same provinces in each year who participated in the optional module on CRC screening. In *Table 6 Part A*, a DD design is employed with the population restricted to that of the pre-intervention group encompassing the CCHS cycles of 2003, 2005 and 2007; the last year serving as the post-intervention period. It is evident from the results for gFOBT screening that very strong and significant effects from the intervention group and the pseudo post-intervention period are present and are fairly similar to the quantities observed in *Table 2*. However, in this test there is no evidence of a significant change in the *Group*Time* interaction term in either the partially controlled (-0.019 [-0.055, 0.017]) or fully controlled (-0.020 [-0.058, 0.015]) sample. Only a significant group effect was observed for the endoscopy outcome in the partially controlled model. The interaction term points to an approximate 2.2 percentage point increase in the procedure which is small but could represent a true effect possibly related to an

uncontrolled province-specific interaction increasing the screening trend. In *Table 6 Part B*, utilizing all five cycles of the CCHS, self-report of having a flu shot was used as the dependent variable. This outcome measure was chosen as a medical procedure open to all individuals but for which there was no policy or other change which coincided with the arrival of the *ColonCancerCheck* screening program in Ontario in 2008. Again, despite displaying strong group and time effects the main effect measure was rejected for significance at the 95% confidence level.

A subsequent bias test in the difference in difference in differences mode was also performed to assess whether there was a change in screening in those under the age of 50, and thus not average risk in absence of family history or other related conditions. If there was an observed increase in this group, it would raise suspicion as to the role of *ColonCancerCheck* given changes observed in a subgroup outside of the mandate of the program. With gFOBT as the dependent variable, results indicated that the change in screening in individuals under the age of 50 in Ontario in the post-intervention period was -0.020 [-0.052, 0.012] percentage points implying no bias of results.

Table 7: Regression Discontinuity Design with Pseudo-Age Cutoffs

Outcome		**Model 1**				**Model 2**			
Past Year Screening		**Mrg Effect**	**95% CI**		**p**	**Mrg Effect**	**95% CI**		**p**
Ontario	**2003-7**	0.0090	-0.0166	0.0346	0.492	-0.0184	-0.0466	0.0098	0.200
	2008-9	0.0190	-0.0217	0.0597	0.361	-0.0352	-0.0833	0.0129	0.152
Control	**2003-7**	-0.0250	-0.0500	-0.0001	0.049	0.0029	-0.0239	0.0297	0.831
	2008-9	0.0091	-0.0245	0.0427	0.594	-0.0049	-0.0358	0.0261	0.758

Using CCHS 2003, 2005, 2007, 2008, 2009
All results are weighted.
Model 1: adjusted for age, ageTh_i, year, province where Th_i =1 if age≥46 and bandwidth is 36-56 years*
Model 2: adjusted for age, ageTh_i, year, province where Th_i =1 if age≥54 and bandwidth is 44-64 years*
Dependent variable is self-report of having a gFOBT in past year or endoscopy (flexible sigmoidoscopy or colonoscopy) in past year as a combined measure. Marginal Effects represented are for treatment variable (Th_i). Results shown are average marginal effects calculated from multivariate logistic regression model
Effects are based on valid responses (complete case analysis used)
Effects based on closest approximation of average risk population available

To examine the validity of the regression discontinuity design I performed several of the same tests as in *Table 4* but altered the treatment variable to reflect a different age. Ages 46 (Model 1) and 54 (Model 2) were selected. These values represent the same degree of departure from the true age threshold and were chosen to avoid being contaminated by

proximity to that effect. Given evidence of sharp random fluctuations in screening by age, particularly in Ontario, examination of pseudo cutoff terms could grant some light as to the robustness of the RDD to these screening trends. For Ontario, the cutoff at age 46 indicates very small and non-significant threshold effect for each time period and the cutoff at age 54 indicates slightly larger, but non-significant negative treatment effects. In the control group in the pre-intervention period in Model 1 there is evidence of a just ($p=0.049$) significant negative effect at the threshold of age 46 at -2.50 [-0.500, -0.00] which appears to be due to a very minor decrease in screening in that area after the cutoff. Although this result appears large enough to constitute concern, the significant findings for post-intervention Ontario in *Table 4* were both much greater in magnitude and robust to several different model parameters including adjustment for covariates and a non-linear age term which increase confidence in those findings displaying true discontinuities.

Discussion

A. Summary and Discussion of Predictors of Screening

This thesis used several methods to measure and explain the effect of *ColonCancerCheck* on screening behavior for prevention of colorectal cancer in asymptomatic average risk adults in Ontario during and after the year 2008. Furthermore, I investigated the predictors of colorectal cancer screening which was in part to highlight covariates which were the most relevant for ensuring exchangeability between treatment groups used in the quasi-experimental models. To my knowledge, this is the first study to measure the effectiveness of *ColonCancerCheck* in a causal framework in terms of its influence on target group-specific screening and the change in screening at the age threshold for becoming average risk for CRC. It is also the first study to investigate how predictors of screening for CRC may be modified upon introduction of such a program.

I found that factors which are consistently associated with screening across group and time categories included advanced age, being physically active, reporting having had a flu shot, having a regular medical doctor and having greater numbers of consultations with

one's doctor. Other predictors of screening behavior were not associated consistently the past year screening outcome or were not very strongly associated. These results are largely in agreement with previous literature on this topic in Canada (19, 23) which is not surprising given that all of these studies use the Canadian Community Health Survey. This project further contributes to this literature for several reasons. Firstly, I compared predictors of screening by intervention group while previously no such distinction was made and findings were attributed to whatever provinces and territories had contributed to the optional module of questions that year. Secondly, I contrasted these groups by intervention period to show how these associations changed over time. In many cases, associations did not consistently hold throughout time and across groups, and the method employed here helps to validate the strength of such associations as well as to allow the opportunity to draw inferences concerning such patterns. Thirdly, I employed greater restrictions on the dataset in order to more closely approximate the asymptomatic average risk population which is the principal focus of most screening programs

Literature reviews commonly find similar forms of variation in the types and strengths of associations as those observed here (97, 98). The importance of the role of a primary care provider has been firmly established in several countries (23, 40, 103, 141) and hypothesized to be related to physicians' capacity to influence beliefs of the value and efficacy of screening for CRC (97). Consistent with this, previous research from Ontario has indicated that nearly one quarter of respondents were ignorant of CRC risks (111) and an additional telephone survey found approximately one half of respondents had not even heard of gFOB tests (112) which strongly implies a shortage of knowledge on the part of many patients. Further evidence suggests that nearly 30% of respondents in a pre-intervention Ontario survey reported no intentions to screen (113), which demonstrates the potential importance of informative physician consultations to amend such beliefs.

Reporting having a flu shot is a related measure which might inform on about the type of individual who is more likely to engage in proactive preventive measures to protect their health. This is consistent with the idea that individuals who visit their doctor more frequently are more likely to screen and similar associations have even been found in regard to dentist visits (97). Alternately, this variable may reflect differences in the

capacity of provincial health care systems in Canada to offer preventive services and engage in large scale campaigns in order to do so. This view is supported in part by vast group-based disparities observed in the proportions of this variable seen in Appendix II Part B, which made this important to control for in subsequent analyses.

Findings also indicate that those who are active are more likely to screen, although there is no difference observed in those who reported themselves as moderately active. This variable might relate again, in part, to the idea of individuals being health conscious. However, if this is so, it is unclear why there are not similarly strong consistent associations observed for variables such smoking status and obesity. It might be possible that being categorized as physically active requires a more determined effort than avoiding cigarettes or obesity. The positive association between screening and advanced age has consistently been observed elsewhere (98) and possibly relates to greater concern with health and knowledge of health risks, in addition to greater familiarity with and use of health care services. In terms of SES, despite there being well documented research elsewhere on its influence on screening variation (97, 98, 100, 108), this study did not derive strong consistent associations based on measures such as education and income. This may be due in part to the available SES measures in the CCHS, how they are used in these analyses, the inclusion of other important confounders, or differences in health care systems or even countries.

B. Summary and Discussion of Total Program Effects

The difference in differences design relied on the suitability of utilizing Ontario as an intervention group and a subset of other Canadian provinces which varied by survey as a control group. The province of Ontario was found to have a greater proportion of individuals living in an urban as opposed to rural setting, higher numbers of immigrants and visible minorities, some evidence of patterns of having higher levels of income and education, and a substantially larger difference in the number of individuals reporting having had a flu shot. A small observed difference in the proportions of individuals with a medical doctor was observed in Ontario and this disparity grew slightly in the post intervention period. There were no consistent differences observed in any of the health and lifestyle variables. While Ontario is clearly not similar to the rest of Canada on all of

these measures, the DD design is only compromised if such differences are changing over time in a manner correlated with the introduction of *ColonCancerCheck*. Such potential changes in associations were examined in Appendix II Part D and further analyzed in difference in difference in differences models.

The difference in difference test was run both with only year and province as covariates, in addition to the main variables, and secondly with a larger covariate set including a number of socio-demographic, lifestyle, and medical variables. The results from the fully controlled version of this test show that *ColonCancerCheck* has increased screening by gFOBT in Ontario in the average risk population by an absolute increase of 5.16 percentage points 95% CI [3.19, 7.12]. The results also indicate that this increase was not observed for endoscopy procedures including optical colonoscopy and flexible sigmoidoscopy at 0.88 [-0.49, 2.25] percentage points. Both the fully controlled and partially controlled models using gFOBT as the dependent variable closely approximated one another, where comparison of these models showed some evidence of confounding by intervention group with endoscopy as the dependent variable. Although there has been some evidence of increases in gFOBT screening use in Ontario following the introduction of *ColonCancerCheck* elsewhere (22) this study is the first to attempt to determine the proportion of this effect which can be plausibly attributed to *ColonCancerCheck* in a method which controls for effects of temporal changes in screening across Canada and time invariant differences in cancer screening within Ontario as a means to remove these biases and specify a true effect.

Given that the gFOBT test is the front line test for CRC screening (22), it is probable that after such a short time since the introduction of *ColonCancerCheck* there would only be an appreciable increase observed in this variety of test. gFOB tests are both administered to a larger segment of the population who do not have an advanced risk designation, and are additionally administered more frequently with biennial recommendation. Due to the high sensitivity and specificity of the endoscopic tests, relatively high performance costs, demanding preparation requirements and health risks associated with perforation and sedation, recommendations suggest wide intervals of screening of up to ten years for colonoscopy and five years for FS (8). It is also likely that even if individuals obtain a

positive gFOB test result, moving them to advanced risk status and justifying endoscopic screening, many will either delay taking the procedure or fail to commit to follow up altogether. Arguably, given the model constraints, gFOBT follow-up may be the most common reason for endoscopic screening. In the Ontario FOBT Pilot Study, median follow-up times following positive gFOBT to endoscopy were 121 days in men and 202 in women (114) and preliminary *ColonCancerCheck* results from 2008 indicate that only 62.1% of individuals obtained a follow-up endoscopy within a six month period (22). It is quite possible that the timeframe available in this study was insufficient to capture a hypothetical increase in endoscopy procedures related to *ColonCancerCheck*.

It is likely that the assumptions of the difference in differences model were not perfectly met. Although there seems to be some evidence of an overall increasing trend towards 2007 in the control group at 12.44% adherence for gFOBT, this trend is much less steep than that seen in Ontario. This is problematic as the DD model assumes that in absence of intervention the trends in the two treatment groups will be parallel (132), and raises the concern of uncontrolled trends in screening which interact with the group and time terms. Temporal trends were partially adjusted for through the inclusion of categorical year variables although this would not ensure full control for this trend. Results from the placebo tests observed in *Table 6 Part A* for the gFOBT outcome variable offer evidence that the supposed measured effect of *ColonCancerCheck* on screening is likely not biased by an uncontrolled trend in screening. When restricting the analysis to the pre-intervention period, no significant effect was found in the *Group*Time* interaction term which would strongly indicate evidence of such a trend. The results from *Table 6 Part A* do suggest, but do not conclude, that there may be some risk of bias of this sort in the endoscopy outcome although this seems less of a concern given the overall null result for that effect and the positive nature of the bias trend. Further evidence of the appropriateness of the model is illustrated in *Part B* of *Table 6*.

A difference in difference in differences model was also run to examine whether the introduction of *ColonCancerCheck* had influenced the associations between certain variables and screening for CRC. In all instances, the variables were observed to increase in their strength of association in Ontario in the post intervention period, possibly

implying that *ColonCancerCheck* was enforcing greater compliance of screening guidelines in groups of people who were habitually more likely to screen to begin with. Although there was an interest in whether *ColonCancerCheck* would increase screening in groups historically less likely to do so under different circumstances, evidence in support of such patterns was not found (Appendix II Part D).

The strongest change in association was related to endoscopic procedures and having a regular medical doctor Ontario post-intervention, indicating a 0.06 [-0.007, 0.137] marginal effect. This does not relate to a greatly increased overall use of endoscopy services in Ontario post-intervention, although could possibly be related to the mode of services delivery alongside it. There was very weak evidence for a change in the same association for the gFOBT outcome which was somewhat unexpected given the central role which physicians were given at this time in the screening program (116). As seen in Appendix II Part F, there was an observed increase in the strength of the association of having a regular doctor interacted with time on gFOBT screening although not by group or the effect measure of interest.

The PCP is most likely either the primary motivator of screening or is visited soon after motivation from a media trigger or other forms of outreach derived from *ColonCancerCheck*. Although there are other paths to obtain gFOBT kits, these are usually restricted to individuals without a PCP (120) which further emphasizes the importance of this association. Additionally, *ColonCancerCheck* aims to educate health care providers and provide them with kits, counseling manuals and other forms of promotional material which could facilitate greater screening adherence in their patients (22). Prior research associated with the pilot program initiated by Cancer Care Ontario found that over 30% physicians were not compliant with recommending screening procedures to age-eligible patients and that many physicians expressed confusion as to the proper protocols involved (20) so this reinforces how such a program element could be beneficial.

Results for the relationship between having a PCP and screening in Appendix II Part D are consistent with the DDD and reinforce that, in both intervention groups, the association of gFOBT screening and having a regular physician increased dramatically in

the post intervention period. However, although this relationship is much larger in Ontario on an absolute scale, comparable relative associations on the odds scale were observed in Ontario (OR 4.13 [2.76, 6.18]) and the control provinces (OR 4.14 [2.64, 6.52]) in the post-intervention period, which seems result from reduced screening in the subset of individuals without a PCP in the control group compared to the pre-intervention period. This may also explain in part how the strength of this association grew without resulting in a larger net increase in screening in the control provinces, although it is not clear whether this association is truly a trend or an unintended result of the unbalanced panel of provinces or even contamination of the intervention outside of Ontario. The DDD indicates a 7.4 percentage point increase in screening among those having a PCP in the post-intervention period (*T*MD*) across Canada which also coincides with this increase. These results suggest that education materials for physicians may not contribute much to the effect of the program if physicians outside of Ontario are similarly becoming more knowledgeable and familiar with such screening tests, assuming similar levels of confusion and noncompliance as in Ontario in earlier periods which does not seem implausible (21).

C. Summary and Discussion of Program Effects at Age Threshold

The regression discontinuity design portion of the analysis was initially conducted for each group and time combination separately and then combined into a DDD style equation to evaluate if there was a treatment effect observed at the age threshold which could be due to *ColonCancerCheck*. As with the DD design, I first compared the distribution of covariates above and below the age threshold term of 50 to examine how close the two treatment groups represented each other within the bandwidth age ranges of 40-60 years and 45-55 years. In many respects the groups appeared to be similar, with perhaps some evidence for greater tendencies for health care use above the age threshold, and the more restrictive bandwidth did not confer sizably greater benefits in group comparability. In *Table 4* I show the results of both partially controlled and fully controlled regression discontinuity designs with age either included as a linear or squared polynomial function. For the gFOBT outcome, the inclusion of additional covariates only slightly attenuated the effect observed in the partially controlled model suggesting fairly

strong exchangeability and minimal confounding as was expected. In all models, a significant threshold effect at 95% confidence was observed only for Ontario in the post intervention period. This magnitude of this effect was slightly diminished in models using a squared age term in comparison to the linear term possibly indicating that, although still fairly similar, a portion of the observed discontinuity at the age of 50 results from a non-linear function of age. No significant effects were observed from the endoscopy outcome. Similarly to the difference in differences analysis, I believe this lack of effect is largely related to the infrequency of the outcome in the average risk group, the greater likelihood for screening at higher ages, and the greater likelihood of a positive gFOBT at higher values of age (22) which might lead to subsequent endoscopy tests. The placebo tests illustrated in *Table 7* provide some evidence that, for the most part, the RDD models are robust and the significant threshold effect is likely large enough to be confidently attributed to a true discontinuity of past year screening with age and not a local aberration in screening trends.

Although encouraging, the results from the regression discontinuity design do not provide proof that *ColonCancerCheck* derived the significant effect of screening in Ontario after the intervention. Due to screening organizations recommending age 50 as the initiation age for screening (6-8) this would likely account for some degree of the threshold effect observed in *Table 4*. Since the RDD cannot differentiate effects of multiple treatments, I controlled for screening adherence at the age threshold in Ontario and the control provinces in order to differentiate program effects from those derived from common baseline screening rates influenced by group and time. *Table 4* and *Figure 2* each illustrate how threshold effects for age both vary by group and change over time, making these important patterns to consider in such a model. As seen in *Table 5*, at a 95% confidence level, none of the results of this design provided significant evidence of an effect on screening at the age threshold for either gFOBT or endoscopy from the screening program. However, partially and fully controlled models examining the gFOB test outcome under a bandwidth of 45-55 years amounted to 4.56 95% CI [-0.40, 9.38] and 4.51 [-0.35, 9.38] percentage point increases, respectively, which are strongly suggestive of a threshold screening effect due to *ColonCancerCheck*. A narrower bandwidth is consistent with the findings of a larger treatment effect if a greater

proportion of the effect is attributed to treatment as opposed to a continuous function of age which might have been the case for the wider bandwidth.

D. Methodological Strengths

Natural or quasi-experiments are generally susceptible to a number of issues concerning validity of findings. Threats most relevant to internal validity include unmeasured variables, trends in outcomes, misspecification of variance, measurement error, endogeneity, selection bias, attrition and omitted interactions. Further problems concerning generalizeability of results may be due to interactions between setting, time, and selection with treatment (142). I will address these concerns in this and the next section in terms of the strengths and suitability of the models used in this project.

The difference in differences design and the differences in difference in differences extension are able to reduce many of these threats to internal validity (142). Use of a control group helps to reduce bias attributable to the differences between healthcare systems which have evolved in somewhat distinctive ways (143) and have demonstrated differences in provision of CRC screening (23). These methods also reduce concern about measurement error, which is often a problem in survey data. The DD models additionally controlled for province of respondent to take into account any disparities within the provinces which constituted the control group. Perhaps most importantly, DD models are able to distinguish a causal effect whilst controlling for certain temporal trends in screening behavior; trends which have been observed to increase screening in several provinces in Canada over time (23). Such temporal trends make it nearly impossible to draw any meaningful inference of program impact from a single group, unless it was somehow possible to establish the absence of such trends.

The regression discontinuity design has a less stringent set of model assumptions than many other natural experiments and is thus thought to be more likely to meet them (138). Due to the ability of this model to approximate randomization of the sample, it is able to avoid many common problems associated with internal validity in natural experiments (144). Additionally, in large part due to the narrow range of focus, it is usually the case that more is known about the selection mechanism and there is less chance for bias by

unmeasured factors (145). This study was not subject to one of the most common concerns relevant to RDD which is that participants can 'self select' or manipulate the treatment variable, thus invalidating the randomization approximation (144), as participants cannot manipulate their age. The RDD alone was not sufficient to measure program-derived effects due to cross-over with existing screening interventions, but did provide a means to draw comparisons over time and across treatment groups, which provided insights for the subsequent causal model.

Use of repeated cross-sectional studies is also beneficial in several respects. An illustrative example from the literature used both cohort and repeated cross-sectional study results to examine screening rates for breast cancer (146). The findings of this study suggested that use of repeated cross sectional data was advantageous in that it did not suffer from cumulative losses to follow-up and better reflected large-scale community level changes such as those due to age demographics. The authors also found that answers on breast cancer screening behavior and attitudes were comparable between the two survey methods (146). This element of study design therefore addresses such internal validity concerns as sample attrition and interactions of the treatment with different settings and times. Additionally, repeated cross sectional data has been noted to help alleviate miscalculations in variance measurement due to correlation of residuals which is often observed in panel data (142). The use of bootstrap repeated replications also provide optimal estimates of true variance for the complex survey design of the CCHS (128).

E. Study Limitations and Methodological Concerns

There are several concerns which should be addressed which relate to the study designs and how they were implemented for the purposes of addressing study questions. The first such concern is the suitability and representativeness of the control group as used in the DD and DDD designs. I have noted previously that these designs are able to remove much of the bias associated with differences between the intervention groups although this does not completely alleviate such concerns. Despite the ability of the DD design to control for time invariant group level differences, it has been demonstrated with comparative DD models that selection of differing control groups influences outcomes. One study demonstrated, using DD models to estimate the impact of US state laws on

changes in log earnings and employment in the construction industry, that the choice of control state greatly influenced the findings even to the extent that the policy interventions were seen as beneficial in some comparisons and detrimental in others (147).

In this study the membership of the control group was not set to a particular province but included all provinces which requested an optional module of CRC screening questions in the CCHS. The inconsistency in year-to-year membership of the control group seemed to distort the temporal trends in screening observed for this group. For instance, in the 2005 year there was a decrease in proportion screened from the previous survey year from 8.34% to 7.03% for the gFOBT outcome which was uncharacteristic of the overall trend. I hypothesize this to be a distortion brought on by the lack of representation from western provinces in 2005 (see *Table 1*). Notably, western provinces have been measured previously as having screening rates close to the national average (23). Although the four Atlantic Provinces only constitute 12% and 10% of the pre and post intervention control group, respectively, they are the most consistently represented and thus heavily influence the trajectory of screening. The presence of the provinces of Manitoba and Alberta only in 2008 is a particular limitation as these provinces are in the process of developing similar programs to *ColonCancerCheck* (121, 122). This is potentially indicative of a greater comparability to Ontario in terms of the infrastructure and resources to produce such services on a mass level, and may influence the magnitude of the time trend and group effect controlled in the DD analysis. Although there is evidence for significant increases in screening between 2005 and 2008 in Newfoundland and Labrador and New Brunswick (23), greater representation from Western provinces may potentially amplify the time trend and make the control group a better counterfactual for Ontario.

A further concern with group selection in difference in differences designs is the presence of unmeasured interactions between group and time variables. This may be due to national or macro level conditions which do not influence provinces equally, or policy changes which differentially impact population subgroups. Furthermore, such unmeasured effects are more likely when intervention groups are less comparable and there are sizable temporal trends (142) which arguably occur in this thesis (see *Table 2*). I have addressed this concern in two ways. First, as noted previously, substantial

differences were observed in the slopes of the intervention groups pre-2008 for both dependent variables which indicates assumptions of the DD model were not perfectly met (132). I demonstrated through placebo tests which limited the timespan to the pre-intervention period and set 2007 as a pseudo-intervention year that, in spite of large period and group effects in gFOBT outcomes, the effect measures were found to be non-significant which is supportive of the DD model being robust to potential unmeasured trends. Secondly, the DDD analysis examined certain variables which were shown to change their associations over time in a pattern consistent with an interactive effect with the group and time variables. Past year screening by gFOBT, observed to increase due to the role of the screening program, was not found to be confounded by any such interactions as observed in *Table 3*. Given the observational nature of this study design there likely remain unmeasured interactions which present the concern of residual confounding. In this project I examined many variables in relation to CRC screening which were chosen due to their commonality in the literature and their face validity, and it seems unlikely that outside of these any unmeasured interaction would be sufficient to have an appreciable distorting effect on the effect outcome.

A final concern in natural experiments with group selection is the problem of endogeneity of policy adoption, sometimes referred to as political economy in the economics literature (142). This concern arises when policymaking is a response to certain conditions which may be political, economic or health related, and distinctive in the province or state in question. Although DD models are able to control for time invariant group differences, they may be insufficient for situations in which policy changes were associated with past province-specific outcomes. That is, the conditions that brought about a policy will have independent effects on the subsequent policy change (147) and the outcome of interest drives policy adoption. An example of this problem is a DD analysis of the effect of a tax reform which used the wives of high and low earning husbands as treatment groups (148). Criticisms of endogeneity have been made in regard to unrealistic assumptions of equal preferences and abilities to work between the groups (149).

Due to guarantees for reasonable access to care, federal oversight of provincial healthcare delivery, harmonized training of health care professionals, and conditional federal transfer

payments to provinces (143), I argue that heterogeneity in health services delivery between the provinces is greatly reduced. Despite this, it is evident that there are disparities in provision of CRC screening services (23) although this may be attributable to patients rather than healthcare systems. Many of the most severe health disparities in Canada occur on north-south lines and reflect access issues related to population density (143). However, these disparities seem to be greatest among the three territories which are excluded from this analysis for such reasons, and urban or rural geography did not seem to greatly influence the outcome within the provinces (See Appendix II Part D).

Although there are various forms of inequity in access to care in Canada by SES, these dynamics vary, but not significantly so, between provinces (150). Age and sex standardized per capita provincial health care expenditures in Ontario amounted to $CAD 2263.8 with the national average at $CAD 2321.4 with most Canadian provinces, but not territories, falling within a few hundred dollars of that range in 2002 (151). Disparities in healthcare quality measures such as unplanned readmission rates for acute myocardial infarction and rates of in-hospital mortality have been shown to be relatively stable with some tendencies to increase in Atlantic Canada (152). Examining wait times for gastroenterologist consultations and procedures indicates that Ontarians, despite higher screening use, tended to wait longer for services than the national median (153).

These findings do not point to differences which I feel are substantial or could drive endogenous policy change. It is conceivable that of all the provinces Ontario obtained a screening program first due to its relative wealth and capacity for innovation, although the fact that two additional provinces are developing similar programs seems to indicate that these differences may not be extensive. The development of the screening program in Ontario also does not seem to be directly linked to a disproportionate disease burden as CRC incidence is much higher in Atlantic Canada (25) and national screening guidelines seem to ensure that the same protocols for screening and treatment are being promoted across Canada (6-8). Additionally, the nature of the dependent variables here avoids many common difficulties in the field of economics with highly auto-correlated dependent variables such as wages and employment which run the risk of understating standard error (154). Here, although it's true that past screening adherence is predictive of present

screening behaviour with the gFOB test although not for FS (98), I feel I can safely assume that this avoids any comparable level of auto-correlation of wages, for instance, which are largely dependent on previously held positions and qualifications (154).

Another study concern relates to the retrospective nature of outcome reporting. The CCHS is collected throughout the year equally from January to December (124) which would mean that if an individual is asked to recall past year screening in 2008 they could easily be speaking to events which occurred in the year 2007 prior to the official launch of *ColonCancerCheck*. A similar effect seems to be illustrated in Appendix III Part B when, if we assume individuals are most likely to screen at age 50, the biggest discontinuities are observed in those aged 51 possibly implying a lagging effect based in retrospective reporting of the outcome. This represents a form of contamination whereby treatment effects, in the province of Ontario, were attributed prior to treatment allocation.

Although this does not fully alleviate this problem, I believe that the March 2008 start date for *ColonCancerCheck* is not entirely accurate. Funding was formally announced in January 2007 (22) and over the next fifteen months, apart from developing screening protocols and building capacities (53, 115), gFOBT kits were sent to over 10 000 physician offices and 3000 pharmacies alongside educational materials and translations of instructions (22). It is not clear exactly when these initiatives took place during these 15 months, but it is likely that interventions linked to *ColonCancerCheck* occurred as early as 2007 and can be seen to lessen this contamination effect. The magnitude of the increase observed in 2008 also seems unlikely to occur under conditions of no intervention and the fact that a comparable and somewhat higher value was observed in 2009, for which contamination is not a concern, provides evidence this was not a one year aberration. There is also a small likelihood of contamination by pilot programs in Ontario (114) and Manitoba (121) although the size and magnitude of these programs were not considered to be large enough to bias the results.

A related concern is the accuracy of the recall of screening outcomes and other variables. A meta-analysis on accuracy of self-reported outcomes for cancer screening measures found that, using a random effects estimate, sensitivity for gFOBT reporting in the past two years was 82% 95% CI [73, 88] with specificity 78% [71, 83] and sensitivity for

screening endoscopy self-report in the past five years was 79% [73, 84] and specificity 90% [85, 93] (155). The authors also noted tendencies of underreporting in CRC screening outcomes (155). Furthermore, recent research in accuracy of administrative databases for the related polypectomy procedure in Quebec also raises suspicion that even administrative records are subject to under-reporting of such outcomes (156). It is possible the use of past year screening outcomes in this thesis may improve accuracy of self-reported screening although there is no way to verify this. Such measurement errors are reduced to an extent by use of control groups (142) and, given the relative consistency of the CCHS in terms of mode of collection and phrasing of questions (131), are also non-differential over time.

As such, it is most likely that the main results presented here in the form of risk differences are downward-biased by imperfect sensitivity and specificity due to non-differential misclassification and it is therefore conceivable that the true effect of *ColonCancerCheck* is larger than reported here. I see no reasons to believe that the accuracy of self-report should change by intervention group in the DD or the RDD given no sizable differences in the covariates which might influence differential misclassification. Although there was a slight tendency for higher SES in Ontario, the literature suggests SES may not greatly influence self-report accuracy (157). In addition, although self-report accuracy decreases at advanced ages, a concern for the RDD, this contrast is typically seen in individuals over 75 (158) who are not included in this study.

Finally, it is likely this study was not fully able to create an average risk prior-to-screening group. Data was only available for family history of CRC and treatment of CRC in regard to questions qualifying screening. Therefore, individuals who did not report *ever* screening with either gFOBT or endoscopy were not subsequently asked their reasons for screening, which were the basis of identification of advanced risk individuals. There is no guarantee that individuals with family histories are more likely to screen with gFOBT but there is strong evidence that many of those with family history or symptoms would have opted for endoscopy (98) although it is still not very clear of how much of the advanced risk population was removed. Again, there is no reason to believe that this bias would be differential by group or time.

F. Implications

This project has come to the conclusion that *ColonCancerCheck* has successfully increased total screening adherence with gFOBT within the time frame studied in the average risk population in addition to providing convincing evidence of an increase in screening at the age threshold for the average risk population at 50 years. This analysis has used an age range and developed a framework for an average risk population consistent with national screening recommendations in Canada (6-8) and the proposed target groups for screening programs in development in Alberta and Manitoba (121, 122). As such, the results are appropriate to the needs of these groups and support the development of such programs and validate the model developed by *ColonCancerCheck*.

There are a few issues with interpretation which should be addressed. Firstly, it is possible that this type of intervention will have a differential impact based on pre-intervention levels of the dependent variables in different provinces (142). This is an important caveat if generalizing the effects of such a program onto a province with a substantially lower pre-intervention screening adherence rate. Secondly, the results of both the gFOBT and endoscopy outcome measures do not easily translate into measures of adherence if policymakers are be interested in effects of a screening program on two year or five year adherence measures. The outcome measures used were chosen to reduce contamination effects derived from misattributing an outcome to the pre-intervention period and may be more reflective of a shift in screening behavior rather than as a direct interpretation of program impact on total screening adherence. Thirdly, I cannot make any claims to have captured the complete effect attributable to *ColonCancerCheck* as this program is currently ongoing and developing new group targeting strategies and other means of increasing screening (22). Future analyses may create a more complete picture of the program's influence. Additionally, this analysis has not been implemented to highlight contributions of specific interventions or strategies within *ColonCancerCheck*. Rather I have focused on capturing evidence of an overall effect making use of causal frameworks for which the specific mechanisms of an intervention are not relevant.

Future research directions could include a follow-up study of *ColonCancerCheck* at a later time period in order to better capture the overall program effect. It would also be of

interest to policymakers to investigate the specific outreach strategies within *ColonCancerCheck* to evaluate the successes of these independently. Finally, measurement of cancer-specific outcomes such as CRC and polyp incidence or measurement of stage of cancer at detection may reflect on an ability of the program to reduce cancer morbidity and mortality through early detection.

Conclusions

The overall impact of the colorectal cancer screening program *ColonCancerCheck* on use of screening services in the asymptomatic average risk population is an important step in assessing the effectiveness of this intervention. Use of natural experiments is a novel approach to this problem which allows greater certainty of the impact of the program while controlling for various trends and group specific effects which may otherwise bias results inferred from examination of screening in the province of Ontario alone.

I first investigated the predictors of colorectal cancer screening to highlight covariates which were the most relevant for ensuring exchangeability between treatment groups used in the quasi-experimental models and to examine any associations which potentially interacted with the group and time terms in such ways which could potentially bias the Difference in Differences results. I found that factors which are consistently associated with screening across group and time categories included advanced age, being physically active, reporting having had a flu shot, having a regular medical doctor and having greater numbers of consultations with one's doctor. These associations were supported in the literature and coincide with previous findings on health consciousness, proactive health care seeking behaviors, and increased awareness and appreciation for colorectal cancer screening.

The Difference in Differences and Difference in Difference in Differences designs utilized Ontario as an intervention group and a subset of other Canadian provinces, which varied by survey year as a control group. The years of 2003, 2005, and 2007 contributed

to a pre-intervention period and 2008-9 for a post-intervention period to align with the formal release of *ColonCancerCheck* in March of 2008. Although there were observed disparities in the characteristics of the participants in Ontario to the rest of Canada, the DD and DDD models controlled for the most relevant of these differences and for the most part I did not observe strong evidence of confounding. The results from the fully controlled test show that *ColonCancerCheck* has increased past-year screening by gFOBT in Ontario in the average risk population by 5.16 percentage points 95% CI [3.19, 7.12]. The results also indicate that this increase was not observed for endoscopy procedures. This method is highly appropriate to measure the impact of a policy which varies at a provincial level in a non-randomized setting over time and removes bias due to the effect of temporal change throughout Canada in CRC screening practices and bias due to time invariant differences in screening between Ontario and other provinces.

A Difference in Difference in Differences model was also run to examine whether the introduction of *ColonCancerCheck* had influenced the associations between certain variables and screening for CRC. This also was a method to verify confounding by unmeasured interactions as part of the DD. No significant findings were found although evidence strongly suggested an increase in the importance of having a regular medical doctor and endoscopies which corresponded to the arrival of *ColonCancerCheck*.

I employed tests to examine if there was a change in screening behavior at the age of 50 which could be due to the influence of *ColonCancerCheck*. Regression Discontinuity Designs were tested on each intervention group in the pre and post intervention periods. Four different models varying by inclusion of covariates and characterization of age, each found strong significant increases only for gFOBT testing in Ontario in the post-period which varied between 7.1 and 9.1 percentage point increases. No significant effects were observed for endoscopy outcomes at the age cutoff. Given the inability of these tests to distinguish a program effect from *ColonCancerCheck* I introduced the RDD threshold treatment term into a DDD model and found convincing evidence of a threshold effect likely attributable to *ColonCancerCheck* for past year screening by gFOBT of 4.51 [-0.35, 9.38] percentage points. The strength of the Regression Discontinuity Design is that it exploits the age-based specification of 50 at which one is considered average risk and

thus a target for increasing screening. Using this policy allows one to create an approximation of randomization with a control group of individuals on one side of the age threshold who are similar in most meaningful respects to the individuals in the treatment group on the other side of the threshold, especially given smaller age bandwidths.

The results from this project bring strong evidence of the effect of *ColonCancerCheck* on increasing screening in the asymptomatic average risk population in Ontario up to the end of 2009. I provide evidence that the proportion of individuals in the average risk population has increased screening and additionally suggest that the program has influenced individuals to screen at greater numbers as they enter the average risk population at age 50. I have validated these results by initiating several tests of model robustness and bias and found no reasons to believe that the results here are biased by unmeasured interactions coinciding with trend in screening or by treatment group disparities leading to endogeneity-derived outcomes. This analysis has used an age range and developed a framework for an asymptomatic average risk population consistent with national screening recommendations in Canada and the proposed target groups for screening programs in development in Alberta and Manitoba. As such, the results support the development of such programs and validate the model for population based colorectal cancer screening of asymptomatic average risk Canadians developed and utilized by *ColonCancerCheck*.

References

1. Canadian Cancer Society. Colorectal Cancer Statistics. Toronto: Canadian Cancer Society; 2011. (http://www.cancer.ca/Canada-wide/About%20cancer/Cancer%20statistics/Stats%20at%20a%20glance/Colorectal%20cancer.aspx?sc_lang=en). (Accessed July 10, 2011).
2. Cancer Care Ontario. Insight on Cancer: News and Information on Colorectal Cancer and Screening in Ontario. Toronto, 2008, (Insight on Cancer)(Canadian Cancer Society (Ontario Division)).
3. CancerCare Manitoba. Cancer in Manitoba Incidence and Mortality 2007 Annual Statistical Report Winnipeg, 2010, (Manitoba Epidemiology and Cancer Registry).
4. Ries LAG, Wingo PA, Miller DS, et al. The annual report to the nation on the status of cancer, 1973–1997, with a special section on colorectal cancer. *Cancer.* 2000;88(10):2398-424.
5. Maroun J, Ng E, Berthelot JM, et al. Lifetime costs of colon and rectal cancer management in Canada. *Chronic Dis Can.* 2003;24(4):91-101.
6. National Committee on Colorectal Cancer Screening. Recommendations for population-based colorectal cancer screening. Ottawa; 2002. (http://www.phac-aspc.gc.ca/publicat/ncccs-cndcc/ccsrec-eng.php). (Accessed June 15, 2011).
7. Canadian Task Force on Preventive Health Care. Colorectal cancer screening: Recommendation statement from the Canadian Task Force on Preventive Health Care. *Canadian Medical Association Journal.* 2001;165(2):206-8.
8. Leddin D, Hunt R, Champion M, et al. Canadian Association of Gastroenterology and the Canadian Digestive Health Foundation: Guidelines on colon cancer screening. *Canadian Journal of Gastroenterology.* 2004;18(2):93-9.
9. Wilson JMG, Jungner G. Principles and practice of screening for disease. Geneva, 1968, (Public Health Papers No. 34)(World Health Organization).
10. Rockey DC. Occult gastrointestinal bleeding. *The New England Journal of Medicine.* 1999;341(1):38-46.

11. Ransohoff DF, Lang CA. CLINICAL GUIDELINE: PART II: Screening for Colorectal Cancer with the Fecal Occult Blood Test: A Background Paper. *Annals of Internal Medicine.* 1997;126(10):811-22.
12. Soares-Weiser K, Burch J, Duffy S, et al. Diagnostic accuracy and cost-effectiveness of faecal occult blood tests (FOBT) used in screening for colorectal cancer: a systematic review. *Centre for Reviews and Dissemination Report*. York: Centre for Reviews and Dissemination, University of York, 2007.
13. Levin B, Lieberman DA, McFarland B, et al. Screening and Surveillance for the Early Detection of Colorectal Cancer and Adenomatous Polyps, 2008: A Joint Guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA: A Cancer Journal for Clinicians.* 2008;58: 130-160. doi: 10.3322/CA.2007.0018
14. Thiis-Evensen E, Hoff GS, Sauar J, et al. Population-based surveillance by colonoscopy: effect on the incidence of colorectal cancer: Telemark Polyp Study I. *Scandinavian Journal of Gastroenterology.* 1999;34(4):414-20.
15. Hoff G, Grotmol T, Skovlund E, et al. Risk of colorectal cancer seven years after flexible sigmoidoscopy screening: randomised controlled trial. *British Medical Journal.* 2009;338:b1846.
16. Atkin WS, Edwards R, Kralj-Hans I, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. *The Lancet.* 2010;375(9726):1624-33.
17. Muller AD, Sonnenberg A. Prevention of colorectal cancer by flexible endoscopy and polypectomy: a case-control study of 32 702 veterans. *Annals of Internal Medicine.* 1995;123(12):904-10.
18. Hewitson P, Glasziou P, Watson E, et al. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. *American Journal of Gastroenterology.* 2008;103(6):1541-9.
19. Sewitch MJ, Fournier C, Ciampi A, et al. Adherence to colorectal cancer screening guidelines in Canada. *BMC Gastroenterol.* 2007;7:E39.
20. Cancer Care Ontario. Ontario FOBT Project Final Report. Toronto, 2006, (Ministry of Health and Long Term Care).

21. Mack LA, Stuart H, Temple WJ. Survey of colorectal cancer screening practices in a large Canadian urban centre. *Can J Surg.* 2004;47(3):189-94.
22. Cancer Care Ontario. Colon Cancer Check 2008 Program Report. Toronto, 2010, (Cancer Care Ontario).
23. Wilkins K, Shields M. Colorectal cancer testing in Canada--2008. *Health Reports.* 2009;20(3):21-30.
24. Andermann A, Blancquaert I, Beauchamp S, et al. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bulletin of the World Health Organization.* 2008;86(4):317-9.
25. Canadian Partnership Against Cancer. The System Performance Initiative: The First Year Report. Toronto, 2009, (The Canadian Partnership Against Cancer).
26. American Joint Committee on Cancer. *American Joint Committee on Cancer Cancer Staging Atlas*. Chicago: Springer-Verlag; 2006.
27. Leslie A, Steele RJC. Management of colorectal cancer. *Postgraduate Medical Journal.* 2002;78(922):473-8.
28. Winawer SJ, O'Brien MJ, Waye JD, et al. Risk and surveillance of individuals with colorectal polyps. WHO Collaborating Centre for the Prevention of Colorectal Cancer. *Bulletin of the World Health Organization.* 1990;68(6):789-95.
29. Morson B. Precancerous conditions of the large bowel. *Proceedings of the Royal Society of Medicine.* 1971;64(9):959-62.
30. Bond J. Clinical evidence for the adenoma-carcinoma sequence, and the management of patients with colorectal adenomas. *Seminars in Gastrointestinal Disease.* 2000;11(4):176-84.
31. Winawer SJ. Natural history of colorectal cancer. *The American Journal of Medicine.* 1999;106(1A):3S-6S.
32. Bond J. Colon polyps and cancer. *Endoscopy.* 2003;35(1):27-35.
33. Kulling D, Christ A, Karaaslan N, et al. Is histological investigation of polyps always necessary? *Endoscopy.* 2001;33(5):428-32.
34. van Dam L, Kuipers EJ, van Leerdam ME. Performance improvements of stool-based screening tests. *Best Practice & Research in Clinical Gastroenterology.* 2010;24(4):479-92.

35. Adamsen S, Kronborg O. Acceptability and compliance in screening for colorectal cancer with fecal occult blood test. *Scandinavian Journal of Gastroenterology.* 1984;19(4):531-4.
36. Hardcastle JD, Chamberlain JO, Robinson MHE, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *The Lancet.* 1996;348(9040):1472-7.
37. Kewenter J, Brevinge H, Engaras B, et al. Results of screening, rescreening, and follow-up in a prospective randomized study for detection of colorectal cancer by fecal occult blood testing: results for 68,308 subjects. *Scandinavian Journal of Gastroenterology.* 1994;29(5):468-73.
38. Kronborg O, Fenger C, Olsen J, et al. Randomised study of screening for colorectal cancer with faecal-occult-blood test. *The Lancet.* 1996;348(9040):1467-71.
39. Weller DP, Owen N, Hiller JE, et al. Colorectal cancer and its prevention: prevalence of beliefs, attitudes, intentions and behaviour. *Australian Journal of Public Health.* 1995;19(1):19-23.
40. Cole SR, Young G, Byrne D, et al. Participation in screening for colorectal cancer based on a faecal occult blood test is improved by endorsement by the primary care practitioner. *Journal of Medical Screening.* 2002;9(4):147-52.
41. Schroy 3rd PC, Lal S, Glick JT, et al. Patient preferences for colorectal cancer screening: how does stool DNA testing fare. *The American Journal of Managed Care.* 2007;13(7):393-400.
42. Zubarik R, Ganguly E, Benway D, et al. Procedure-related abdominal discomfort in patients undergoing colorectal cancer screening: a comparison of colonoscopy and flexible sigmoidoscopy. *The American Journal of Gastroenterology.* 2002;97(12):3056-61.
43. Scott RG, Edwards JT, Fritschi L, et al. Community-Based Screening by Colonoscopy or Computed Tomographic Colonography in Asymptomatic Average-Risk Subjects. *The American College of Gastroenterology.* 2004;99(6):1145-51.

44. Coombs A, Jones-McLean E, Le-Petit C, et al. Technical Report for the National Commitee on Colorectal Cancer Screening. Ottawa, 2002, (Health Canada).
45. Lansdorp-Vogelaar I, Knudsen AB, Brenner H. Cost-effectiveness of colorectal cancer screening--an overview. *Best Practice & Research.* 2010;Clinical Anaesthesiology. 24(4):439-49.
46. Pignone M, Saha S, Hoerger T, et al. Cost-Effectiveness Analyses of Colorectal Cancer Screening. *Annals of Internal Medicine.* 2002;137(2):96-104.
47. Whynes DK, Nottingham FOB Screening Trial. Cost-effectiveness of screening for colorectal cancer: evidence from the Nottingham faecal occult blood trial. *Journal of Medical Screening.* 2004;11(1):11-5.
48. Flanagan WM, Le Petit C, Berthelot J-M, et al. Potential impact of population-based colorectal cancer screening in Canada. *Chronic Dis Can.* 2003;24(4):81-8.
49. Ahlquist DA, McGill DB, Fleming JL, et al. Patterns of occult bleeding in asymptomatic colorectal cancer. *Cancer.* 1989;63(9):1826-30.
50. Herzog P, Holtermuller KH, Preiss J, et al. Fecal blood loss in patients with colonic polyps: a comparison of measurements with 51chromium-labeled erythrocytes and with the Haemoccult test. *Gastroenterology.* 1982;83(5):957-62.
51. Young GP, John DJBS, Winawer SJ, et al. Choice of fecal occult blood tests for colorectal cancer screening: recommendations based on performance characteristics in population studies. *American Journal of Gastroenterology.* 2002;97(10):2499-507.
52. Pignone M, Campbell MK, Carr C, et al. Meta-analysis of dietary restriction during fecal occult blood testing. *Effective Clinical Practice.* 2001;4(4):150-6.
53. Rabeneck L, Zwaal C, Goodman JH, et al. Cancer Care Ontario guaiac fecal occult blood test (FOBT) laboratory standards: Evidentiary base and recommendations. *Clinical Biochemistry.* 2008;41(16-17):1289-305.
54. Kahi CJ, Imperiale TF. Do aspirin and nonsteroidal anti-inflammatory drugs cause false-positive fecal occult blood test results? A prospective study in a cohort of veterans. *The American Journal of Medicine.* 2004;117(11):837-41.
55. Pye G, Ballantyne KC, Armitage NC, et al. Influence of non-steroidal anti-inflammatory drugs on the outcome of faecal occult blood tests in screening for

colorectal cancer. *British Medical Journal (Clinical Res Ed).* 1987;294(6586):1510-1.

56. Clarke P, Jack F, Carey FA, et al. Medications with anticoagulant properties increase the likelihood of a negative colonoscopy in faecal occult blood test population screening. *Colorectal Disease.* 2006;8(5):389-92.
57. Li S, Wang H, Hu J, et al. New immunochemical fecal occult blood test with two-consecutive stool sample testing is a cost-effective approach for colon cancer screening: Results of a prospective multicenter study in Chinese patients. *International Journal of Cancer.* 2006;118(12):3078-83.
58. Mandel JS, Church TR, Bond JH, et al. The Effect of Fecal Occult-Blood Screening on the Incidence of Colorectal Cancer. *New England Journal of Medicine.* 2000;343(22):1603-7.
59. Rennert G, Rennert HS, Miron E, et al. Population Colorectal Cancer Screening with Fecal Occult Blood Test. *Cancer Epidemiology Biomarkers & Prevention.* 2001;10(11):1165-8.
60. Michalek AM, Cummings KM, Gamble D. The use of a cancer registry in a mass screening program for colorectal cancer. *Cancer Detect Prev.* 1988;11(3-6):353-7.
61. Winawer SJ, Schottendeld D, Miller D, et al. Detection of early colon cancer and colonic polyps. In: Neiburgs HE, ed. *Prevention and Detection of Cancer: Proceedings from the Third International Symposium on Detection and Prevention of Cancer, New York 1976, Part II, Vol 2*. New York: Marcel Dekker, 1980:2103-10.
62. Sung JJY, Chan FKL, Leung WK, et al. Screening for colorectal cancer in Chinese: Comparison of fecal occult blood test, flexible sigmoidoscopy, and colonoscopy. *Gastroenterology.* 2003;124(3):608-14.
63. Collins JF, Lieberman DA, Durbin TE, et al. Accuracy of Screening for Fecal Occult Blood on a Single Stool Sample Obtained by Digital Rectal Examination: A Comparison with Recommended Sampling Practice. *Annals of Internal Medicine.* 2005;142(2):81-5.
64. Bretthauer M. Evidence for colorectal cancer screening. *Best Practice & Research in Clinical Gastroenterology.* 2010;24(4):417-25.

65. Levin TR, Farraye FA, Schoen RE, et al. Quality in the technical performance of screening flexible sigmoidoscopy: recommendations of an international multi-society task group. *Gut.* 2005;54(6):807-13.
66. Imperiale TF, Wagner DR, Lin CY, et al. Risk of advanced proximal neoplasms in asymptomatic adults according to the distal colorectal findings. *New England Journal of Medicine.* 2000;343(3):169-74.
67. Atkin W, Rogers P, Cardwell C, et al. Wide variation in adenoma detection rates at screening flexible sigmoidoscopy* 1. *Gastroenterology.* 2004;126(5):1247-56.
68. Bretthauer M, Skovlund E, Grotmol T, et al. Inter-endoscopist variation in polyp and neoplasia pick-up rates in flexible sigmoidoscopy screening for colorectal cancer. *Scandinavian journal of gastroenterology.* 2003;38(12):1268-74.
69. Fracchia M, Senore C, Armaroli P, et al. Assessment of the multiple components of the variability in the adenoma detection rate in sigmoidoscopy screening, and lessons for training. *Endoscopy.* 2010;42(EPub):448-55.
70. Selby JV, Friedman GD, Quesenberry Jr CP, et al. A case–control study of screening sigmoidoscopy and mortality from colorectal cancer. *New England Journal of Medicine.* 1992;326(10):653-7.
71. Newcomb PA, Norfleet RG, Storer BE, et al. Screening sigmoidoscopy and colorectal cancer mortality. *Journal of the National Cancer Institute.* 1992;84(20):1572-5.
72. Newcomb PA, Storer BE, Morimoto LM, et al. Long-term efficacy of sigmoidoscopy in the reduction of colorectal cancer incidence. *Journal of the National Cancer Institute.* 2003;95(8):622-5.
73. Hoff G, Dominitz JA. Contrasting US and European approaches to colorectal cancer screening: which is best? *Gut.* 2010;59(3):407-14.
74. Gatto NM, Frucht H, Sundararajan V, et al. Risk of Perforation After Colonoscopy and Sigmoidoscopy: A Population-Based Study. *Journal of the National Cancer Institute.* 2003;95(3):230-6.
75. Anderson ML, Pasha TM, Leighton JA. Endoscopic perforation of the colon: lessons from a 10-year study. The *American Journal of Gastroenterology.* 2000;95(12):3418-22.

76. Frühmorgen P, Demling L. Complications of Diagnostic and Therapeutic Colonoscopy in the Federal Republic of Germany. Results of an Inquiry Komplikationen der diagnostischen und therapeutischen Koloskopie in der Bundesrepublik Deutschland-Ergebnisse einer Umfrage. *Endoscopy.* 1979;11(2):146-50.
77. Tran DQ, Rosen L, Kim R, et al. Actual colonoscopy: What are the risks of perforation? Discussion. *The American surgeon.* 2001;67(9):845-8.
78. Levin TR, Zhao W, Conell C, et al. Complications of Colonoscopy in an Integrated Health Care Delivery System. *Annals of Internal Medicine.* 2006;145(12):880-6.
79. Rex DK, Bond JH, Winawer S, et al. Quality in the technical performance of colonoscopy and the continuous quality improvement process for colonoscopy: recommendations of the U. S. Multi-Society Task Force on Colorectal Cancer. *The American Journal of Gastroenterology.* 2002;97(6):1296-308.
80. Pickhardt PJ, Nugent PA, Mysliwiec PA, et al. Location of Adenomas Missed by Optical Colonoscopy. *Annals of Internal Medicine.* 2004;141(5):352-9.
81. Robertson DJ, Greenberg ER, Beach M, et al. Colorectal cancer in patients under close colonoscopic surveillance. *Gastroenterology.* 2005;129(1):34-41.
82. Winawer SJ, Zauber AG, Ho MN, et al. Prevention of colorectal cancer by colonoscopic polypectomy. *New England Journal of Medicine.* 1993;329(27):1977-81.
83. Citarda F, Tomaselli G, Capocaccia R, et al. Efficacy in standard clinical practice of colonoscopic polypectomy in reducing colorectal cancer incidence. *Gut.* 2001;48(6):812-5.
84. Gross CP, Andersen MS, Krumholz HM, et al. Relation Between Medicare Screening Reimbursement and Stage at Diagnosis for Older Patients With Colon Cancer. *JAMA: The Journal of the American Medical Association.* 2006;296(23):2815-22.
85. Rabeneck L, Paszat LF, Saskin R, et al. Association between colonoscopy rates and colorectal cancer mortality. *American Journal of Gastroenterology.* 2010;105(7):1627-32.

86. Selby JV, Friedman GD, Quesenberry J, Charles P., et al. Effect of Fecal Occult Blood Testing on Mortality from Colorectal Cancer: A Case–Control Study. *Annals of Internal Medicine.* 1993;118(1):1-6.
87. Wahrendorf J, Robra B-P, Wiebelt H, et al. Effectiveness of colorectal cancer screening: results from a population-based case-control evaluation in Saarland, Germany. *European Journal of Cancer Prevention.* 1993;2(3):221-8.
88. Scholefield JH, Moss S, Sufi F, et al. Effect of faecal occult blood screening on mortality from colorectal cancer: results from a randomised controlled trial. *Gut.* 2002;50(6):840-4.
89. Kronborg O, Jørgensen O, Fenger C, et al. Randomized study of biennial screening with a faecal occult blood test: results after nine screening rounds. *Scandinavian Journal of Gastroenterology.* 2004;39(9):846-51.
90. Mandel JS, Bond JH, Church TR, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. *New England Journal of Medicine.* 1993;328(19):1365-71.
91. Mandel JS, Church TR, Ederer F, et al. Colorectal Cancer Mortality: Effectiveness of Biennial Screening for Fecal Occult Blood. *Journal of the National Cancer Institute.* 1999;91(5):434-7.
92. Faivre J, Dancourt V, Lejeune C, et al. Reduction in colorectal cancer mortality by fecal occult blood screening in a French controlled study. *Gastroenterology.* 2004;126(7):1674-80.
93. Taylor C, Schultz SE, Paszat LF, et al. Prevalence of screening in patients newly diagnosed with colorectal cancer in Ontario. *Canadian Journal of Gastroenterology.* 2007;21(12):805-8.
94. Bressler B, Lo C, Amar J, et al. Prospective evaluation of screening colonoscopy: who is being screened? *Gastrointest Endosc.* 2004;60(6):921-6.
95. Raza M, Bernstein CN, Ilnyckyj A. Canadian physicians' choices for their own colon cancer screening. *Canadian Journal of Gastroenterology.* 2006;20(4):281-4.
96. Strumpf EC, Chai Z, Kadiyala S. Adherence to cancer screening guidelines across Canadian provinces: an observational study. *BMC Cancer.* 2010;10(1):E304.

97. Vernon SW. Participation in Colorectal Cancer Screening: A Review. *Journal of the National Cancer Institute.* 1997;89(19):1406-22.
98. Jepson R, Clegg A, Forbes C, et al. The determinants of screening uptake and interventions for increasing uptake: a systematic review. *Health Technology Assessment* 2000;4(14):1-133.
99. Pruitt SL, Shim MJ, Mullen PD, et al. Association of area socioeconomic status and breast, cervical, and colorectal cancer screening: a systematic review. *Cancer Epidemiology, Biomarkers & Prevention.* 2009;18(10):2579-99.
100. von Wagner C, Good A, Wright D, et al. Inequalities in colorectal cancer screening participation in the first round of the national screening programme in England. *British Journal of Cancer.* 2009;101(S2):S60-S3.
101. Klabunde CN, Riley GF, Mandelson MT, et al. Health plan policies and programs for colorectal cancer screening: a national profile. *American Journal of Managed Care.* 2004;10(4):273-9.
102. Federici A, Rossi P, Bartolozzi F, et al. The Role of GPs in Increasing Compliance to Colorectal Cancer Screening: A Randomised Controlled Trial (Italy). *Cancer Causes and Control.* 2006;17(1):45-52.
103. Vinker S, Nakar S, Rosenberg E, et al. The role of family physicians in increasing annual fecal occult blood test screening coverage: a prospective intervention study. *Israel Medical Association Journal.* 2002;4(6):424-5.
104. Myers RE, Vernon SW, Tilley BC, et al. Intention to screen for colorectal cancer among white male employees. *Preventive medicine.* 1998;27(2):279-87.
105. Frew E, Wolstenholme J, Whynes D. Mass population screening for colorectal cancer: factors influencing subjects' choice of screening test. *Journal of Health Services Research & Policy.* 2001;6(2):85-91.
106. Friedman LC, Everett TE, Peterson L, et al. Compliance with fecal occult blood test screening among low-income medical outpatients: a randomized controlled trial using a videotaped intervention. *Journal of Cancer Education.* 2001;16(2):85-8.

107. Giorgi Rossi P, Federici A, Bartolozzi F, et al. Trying to improve the compliance to colorectal cancer screening: a complex study design for a complex planning question. *Controlled Clinical Trials.* 2005;26:323-30.
108. Wardle J, McCaffery K, Nadel M, et al. Socioeconomic differences in cancer screening participation: comparing cognitive and psychosocial explanations. *Social Science & Medicine.* 2004;59(2):249-61.
109. Wee CC, McCarthy EP, Phillips RS. Factors associated with colon cancer screening: the role of patient factors and physician counseling. *Preventive Medicine.* 2005;41(1):23-9.
110. Klabunde CN, Schenck AP, Davis WW. Barriers to Colorectal Cancer Screening Among Medicare Consumers. *American Journal of Preventive Medicine.* 2006;30(4):313-9.
111. Hoffman-Goetz L, Thomson MD, Donelle L. Reasons for declining colorectal cancer screening by older Canadians: a pilot study. *Journal of Cancer Education.* 2008;23(1):32-6.
112. Ritvo P, Myers R, Del Giudice ME, et al. Fecal occult blood testing: people in Ontario are unaware of it and not ready for it. *Can Fam Physician.* 2009;55(2):176-7.
113. Marshall DA, Johnson FR, Phillips KA, et al. Measuring patient preferences for colorectal cancer screening using a choice-format survey. *Value Health.* 2007;10(5):415-30.
114. Paszat L, Rabeneck L, Kiefer L, et al. Endoscopic follow-up of positive fecal occult blood testing in the Ontario FOBT Project. *Canadian Journal of Gastroenterology.* 2007;21(6):379-82.
115. Rabeneck L, Rumble RB, Axler J, et al. Cancer Care Ontario colonoscopy standards: standards and evidentiary base. *Canadian Journal of Gastroenterology.* 2007;21(Suppl D):5D-24D.
116. Guidice LD, Meuser J. Clinical Tool: Screening Evidence Summary. Toronto: Queens Printer for Ontario; 2008. (http://www.health.gov.on.ca/en/pro/programs/coloncancercheck/screeningresearch.aspx). (Accessed June 15, 2011).

117. Moss S, Ancell-Park R, Brenner H. Evaluation and Interpretation of Screening Outcomes. In: Patnick J, Segnan N, von Karsa L, eds. *European Guidelines for Quality Assurance in Colorectal Cancer Screening and Diagnosis First Edition*. Lyon: Publications Office of the European Union, 2010:71-102.
118. Immunostics Inc. Hema Screen. Ocean, NJ, 2004.
119. U. K. Colorectal Cancer Screening Pilot Group. Results of the first round of a demonstration pilot of screening for colorectal cancer in the United Kingdom. *British Medical Journal.* 2004;329(7458):133-5.
120. Ontario Ministry of Health and Long Term Care. Health Care Professionals: Frequently Asked Questions -Pharmacists. Toronto: Queen's Printer for Ontario; 2011.(http://www.health.gov.on.ca/en/pro/programs/coloncancercheck/pharmacists_faq.aspx). (Accessed June 13, 2010).
121. ColonCheck Manitoba. Phase 1 Summary of Results. Winnipeg: CancerCare Manitoba; 2010. (http://www.cancercare.mb.ca/home/prevention_and_screening/professional_screening_programs/colorectal_cancer_screening/). (Accessed Aug. 1, 2011).
122. Alberta Health Services. Alberta Colorectal Cancer Screening Program. Edmonton: MyHealthAlberta.ca; 2011. (http://www.albertahealthservices.ca/services.asp?pid=service&rid=1026405). (Accessed July 15, 2011).
123. Alberta Health Services. ACRCSP Bulletin. Edmonton: Alberta Health Services; 2010. (http://www.albertahealthservices.ca/1781.asp). (Accessed Aug 1, 2011).
124. Statistics Canada. Canadian Community Health Survey (CCHS)- Annual Component: User Guide 2009 Microdata files Ottawa, June 2010, (Statistics Canada).
125. StataCorp. Stata: release 11. StataCorp LP, 2009.
126. Groves RM, Fowler FJ, Couper MP, et al. *Survey Methodology*. Hoboken, NJ: John Wiley & Sons Inc.; 2004.
127. Lehtonen R, Pahkinen E. *Practical Methods for Design and Analysis of Complex Surveys 2nd Ed.* Chichester: John Wiley & Sons, Ltd.; 2004.

128. Statistics Canada. Canadian Community Health Survey (CCHS). Ottawa; 2010. (http://www.statcan.gc.ca/cgi-bin/imdb/p2SV.pl?Function=getSurvey&SDDS=3226&lang=en&db=imdb&adm=8&dis=2). (Accessed Sept 1 2010).
129. Korn EL, Graubard BI. *Analysis of Health Surveys*. New York: John Wiley & Sons, Inc.; 1999.
130. Thomas S, Tremblay S. Interpreting Estimates from the Redesigned Canadian Community Health Survey (CCHS). Ottawa: Statistics Canada; 2010. (http://www.statcan.gc.ca/imdb-bmdi/document/3226_D64_T9_V1-eng.htm). (Accessed July 15, 2011).
131. Thomas S, Wannell B. Combining cycles of the Canadian Community Health Survey. *Health Rep.* 2009;20(1):53-8.
132. Angrist JD, Pischke JS. *Mostly Harmless Econometrics: an Empiricist's Companion*. Princeton: Princeton University Press; 2009.
133. Card D, Krueger AB. Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *The American Economic Review.* 1994;84(4):772-93.
134. Hollenbeak CS, Gorton CP, Tabak YP, et al. Reductions in mortality associated with intensive public reporting of hospital outcomes. *American Journal of Medical Quality.* 2008;23(4):279-86.
135. Lairson DR, Yoon SJ, Carter PM, et al. Economic evaluation of an intensified disease management system for patients with type 2 diabetes. *Disease Management.* 2008;11(2):79-94.
136. Acs G, Nelson S. Changes in living arrangements during the late 1990s: Do welfare policies matter? *Journal of Policy Analysis and Management.* 2004;23(2):273-90.
137. Strumpf E. Medicaid's effect on single women's labor supply: Evidence from the introduction of Medicaid. *Journal of Health Economics.* 2011;30(3):531-48.
138. Hahn J, Todd, P., Van der Klaauw, W. Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica.* 2001;69(1):201-9.

139. Carpenter C, Dobkin C. The Minimum Legal Drinking Age and Public Health. *The Journal of Economic Perspectives.* 2011;25(2):133-56.
140. Duflo E. Empirical Methods: MIT 14.771/Harvard 2390b. 2002. (http://web.mit.edu/14.771/www/emp_handout.pdf). (Accessed July 15, 2010).
141. Federici A, Giorgi Rossi P, Bartolozzi F, et al. The role of GPs in increasing compliance to colorectal cancer screening: a randomised controlled trial. *Cancer Causes and Control.* 2006;17:45 - 52.
142. Meyer BD. Natural and Quasi-Experiments in Economics. *Journal of Business & Economic Statistics.* 1995;13(2):151-61.
143. Romanow RJ. Building on Values: the Future of Health Care in Canada. Ottawa: Commission on the Future of Health Care in Canada, 2002, (Privy Council publication no. 0660189399)
144. Lee DS, Lemieux T. Regression Discontinuity Designs in Economics. *Journal of Economic Literature.* 2010;48:281-355.
145. Shadish WR, Luellen JK. Regression Discontinuity Design. In: Everitt BS, Howell DC, eds. *Encyclopedia of Statistics in Behavioral Science*. Chichester: John Wiley & Sons, Ltd, 2005:1725-7.
146. Caplan LS, Lane DS, Grimson R. The use of cohort vs repeated cross-sectional sample survey data in monitoring changing breast cancer screening practices. *Preventive Medicine.* 1995;24(6):553-6.
147. Besley T, Case A. Unnatural experiments: estimating the incidence of endogenous policies. *Economic Journal.* 2000;110(467):F672-F94.
148. Eissa N, Heckman, J.J.(Comment). Labor supply and the economic recovery Tax Act of 1981. In: Feldstein M, Poterba, J., ed. *Empirical Foundations of Household Taxation*: University Of Chicago Press, 1996:5-38.
149. Blundell R, Duncan A, Meghir C. Estimating Labor Supply Responses Using Tax Reforms. *Econometrica.* 1998;66(4):827-61.
150. Allin S. Does equity in healthcare use vary across Canadian provinces. *Health Policy.* 2008;3(4):83-99.

151. Canadian Institute for Health Information. Provincial and Territorial Government Health Expenditure by Age Group, Sex and Major Category: Recent and Future Growth Rates. Ottawa: CIHI, 2005, (Canadian Institute of Health Information).
152. Canadian Institute for Health Information. Health Indicators. Ottawa: CIHI, 2008, (Canadian Institute of Health Information).
153. Armstrong D, Barkun ANG, Chen Y, et al. Access to specialist gastroenterology care in Canada: The Practice Audit in Gastroenterology (PAGE) wait times program. *Canadian Journal of Gastroenterology.* 2008;22(2):155-60.
154. Bertrand M, Duflo E, Mullainathan S. How Much Should We Trust Differences-in-Differences Estimates?*. *Quarterly Journal of Economics.* 2004;119(1):249-75.
155. Rauscher GH, Johnson TP, Cho YI, et al. Accuracy of Self-Reported Cancer-Screening Histories: A Meta-analysis. *Cancer Epidemiology Biomarkers & Prevention.* 2008;17(4):748-57.
156. Wyse JM, Joseph L, Barkun AN, et al. Accuracy of administrative claims data for polypectomy. [published online ahead of print June 13, 2011]. *Canadian Medical Association Journal*, 2011:(doi:10.1503/cmaj.100897).
157. Reijneveld SA, Stronks K. The validity of self-reported use of health care across socioeconomic strata: a comparison of survey and registration data. *International Journal of Epidemiology.* 2001;30(6):1407-14.
158. Raina P, Torrance-Rynard V, Wong M, et al. Agreement between Self-reported and Routinely Collected Health-care Utilization Data among Seniors. *Health Services Research.* 2002;37(3):751-74.
159. Rust KF, Rao JNK. Variance Estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research.* 1996;5:283-310.
160. Chowhan J, Buckley NJ. Using mean bootstrap weights in stata: A BSWREG revision. *Statistics Canada Research Data Centres:Information and Technical Bulletin.* 2005;2(1):23-38.
161. Efron B, Tibshirani R. *An introduction to the bootstrap*. New York: Chapman and Hall; 1993.
162. Rao JNK, Wu CFJ. Resampling Inference with complex survey data. *Journal of the American Statistical Association.* 1988;83(401):231-41.

163. Lahiri P. On the impact of survey sampling and small area estimation. *Statistical Science.* 2003;18(2):199-210.

164. Zhang F, Brick M, Kaufman S, et al. Variance estimation of imputed survey data. Washington, DC, 1998, (Working Paper No. 98-14,)(U.S. Department of Education, Office of Educational Research and Development, National Center for Education Statistics).

165. Pfeffermann D, Tiller R. Bootstrap approximation to prediction MSE for state-space models with estimated parametres. *Journal of Time Series Analysis.* 2002;26(6):893-916.

Appendix I

A. Survey Variance

To illustrate the inability of straightforward forms of variance calculation to obtain accurate estimates from a complex survey design I will briefly describe variance estimation and survey design and weighting. The sampling variance estimate of a mean value of *y* for a simple random sample can be represented by the following equation:

$$v(\bar{y}) = \frac{(1-f)}{n} s^2$$

Where: Where: *f* is the sampling fraction=n/N and *(1-f)* is the finite population correction, or *fpc* which represents the proportion of frame elements not sampled. The *fpc* is often ignored when *f* is small and *fpc* approximates 1.

In clustered samples, the mean and sampling variance of the mean is given as follows:

$$\bar{y} = \frac{\sum_{\alpha=1}^{a} \sum_{b=1}^{B} y_{\alpha\beta}}{aB} \qquad s_a^2 = \left(\frac{f}{a-1}\right) \sum_{\alpha=1}^{a} (\bar{y}_\alpha - \bar{y})^2$$

Where: α=1,2,…*a* clusters; β=1,2,….*B* subjects in each cluster

The randomization of the procedure is applied only to the clusters. Clustering will tend to increase the standard error of the mean from what it would be as a simple random sample. This is a result of greater homogeneity within clusters, or intra cluster homogeneity.

In stratified samples, the mean and variance can be expressed as:

$$\bar{y}_{st} = \sum_{h=1}^{H} W_h \bar{y}_h$$

$$v(\bar{y}_h) = \left(\frac{1 - f_h}{n_h} \right) s_h^2$$

$$s_h^2 = \left(\frac{1}{n_h - 1} \right) \sum_{i=1}^{n_h} (\bar{y}_{hi} - \bar{y}_h)^2$$

$$v(\bar{y}_{st}) = \sum_{h=1}^{H} W_h^2 \left(\frac{1 - f_h}{n_h} \right) s_h^2$$

Where: W_h is stratum specific weights for population; h denotes a stratum, $\bar{y}_h$ is stratum specific mean.

B. Bootstrap Re-Sampling Technique

Here, I will describe the underlying process and rationale for using bootstrap repeated replication techniques. In complex, multi-stage designs, sample reuse methods can be used as a means to approximate variance for nonlinear estimators $\hat{\theta}$ which may include ratios and coefficients. Reuse methods take a single sample from the population and estimate variance through a repeated utilization of this sample. Reuse methods include balanced half samples, jackknife, and bootstrap techniques, and follow the general format:

4. K pseudo-samples are drawn from the data set
5. An estimate $\hat{\theta}_k$ mimicking the parent estimator $\hat{\theta}$ is drawn from each of the pseudosamples
6. V($\hat{\theta}$) of the estimator $\hat{\theta}$ is estimated by using the observed variation of $\hat{\theta}_k$ as in traditional variance estimates, based on $(\hat{\theta}_k - \hat{\theta})^2$.

(127)

The bootstrap technique for variance estimation is different from the other methods mentioned primarily in the sampling of pseudo-samples from the original population. The method is to take a sample of clusters (≥2) from the strata of the original sample with replacement and to take a simple random sample of individuals from this group by strata. This is repeated K times to form K independent bootstrap samples (127). The variance estimate typically is as follows:

$$v_b(\hat{\theta}) = \left(\frac{1}{B}\right)\sum_{b=1}^{B}(\hat{\theta}_b - \hat{\theta})^2$$

Where:

$$\hat{\theta} = \left(\frac{1}{B}\right)\sum_{b=1}^{B}\hat{\theta}_b$$

$$\hat{\theta}_b = \sum_{h=1}^{H}\sum_{i=1}^{n_h}\sum_{j\in i} w_{hij}^{b} x_{hij}$$

Where: there are b=1….B independent samples with replacement; $\hat{\theta}_b$ estimator for a given replicate, and $\hat{\theta}$ is the replicate mean. *h*=1…*H* strata, *nh* 1….*I* PSUs, and person *j*, where all persons j are part of PSU *i*. *x* is a given variable with associated replicate specific weights w^b which are specific for a given person, PSU and stratum (159).

Often, each sample is drawn to equal size of the number of units in the dataset. Weights are typically assigned using equivalent cluster and multi-stage designs that produced the final weights of the sample. In selecting units for replicates, selected units receive positive bootstrap weights and weights of zero are assigned to units not selected. The number of bootstrap replications is meant to derive a sample which is consistent, and is often chosen specifically for the survey in question, and increased numbers of replicates can increase variance precision (160).

It has been argued that the bootstrap technique can be used to enlarge the types of statistical problems which can be analyzed, reduces the number of assumptions required to validate analyses, and helps reduce the effort in assessing accuracy (161). It has been demonstrated that nominal error rates in nonlinear statistics, like ratios and regression coefficients, could be more accurately measured with bootstrap resampling methods than with similar measures based on normal approximations (162). It has also been demonstrated that, under conditions of unknown MSE($\hat{\theta}$) the bootstrap technique can capture both upper and lower error rates better than the linearization and jackknife procedures, but is equivalent to these two in calculating the total error rate. (162). Bootstrap variance has shown to be consistently accurate with both smooth/linear

statistics and non-smooth/categorical statistics (159) . Bootstrap replication techniques have been used successfully for studies with complex designs and finite populations. Lahari (163) cites examples of the National Centre of Education Statistics where bootstrap methods were shown to give more accurate standard error estimates which would have been underestimated otherwise especially for variables with high levels of imputation (164), and the US bureau of Labour Statistics where the bias-corrected bootstrap estimator was able to decrease bias associated with MSE (165). Lehtonen and Pahkinen (127) took the example of the Mini-Finland Health Survey which followed a two-stage stratified cluster design and compared sample reuse methods (bootstrap, BRR, jackknife) and linearization in estimating smooth and non-smooth statistics on systolic blood pressure and chronic morbidity, respectively. The authors examined the design effect of each of the techniques and found that there were no significant differences between the methods in terms of their variance calculations.

Appendix II

A. Provincial Composition in Pre and Post Intervention Periods

Province	Pre Intvn	Post Intvn	Total
Newfoundland and Labrador	0.0507	0.0227	0.0377
Prince Edward Island	0.0073	0.0059	0.0067
Nova Scotia	0.0346	0.0415	0.0378
New Brunswick	0.0280	0.0332	0.0304
Quebec	0.0000	0.1790	0.0830
Ontario	0.7079	0.5066	0.6145
Manitoba	0.0000	0.0218	0.0101
Saskatchewan	0.0340	0.0370	0.0354
Alberta	0.0000	0.0642	0.0298
British Columbia	0.1375	0.0881	0.1146

Using CCHS 2003, 2005, 2007, 2008, 2009
*N=55444, Weighted N=*2739322
All proportions are weighted
Proportions are based on valid responses (complete case analysis used)
Proportions based on closest approximation of average risk population available ages 50-74

B. Intervention and Control Group Comparisons for DD and DDD Analyses

Variable	Pre Intervention			Post Intervention		
	Control Prov	Ontario	p	Control Prov	Ontario	p
Marital Status						
Married	0.7431	0.7299	0.067	0.6565	0.7227	0.000
Common-Law	0.0407	0.0469	0.082	0.0781	0.0481	0.000
Widowed	0.0656	0.0662	0.848	0.0611	0.0593	0.673
Separated	0.0259	0.0274	0.560	0.0280	0.0322	0.253
Divorced	0.0727	0.0760	0.395	0.0961	0.0776	0.004
Single, Never Married	0.0520	0.0536	0.614	0.0803	0.0600	0.000
Country of Birth						
Canada	0.8163	0.6536	0.000	0.8420	0.6052	0.000
Other N .America	0.0165	0.0155	0.611	0.0136	0.0157	0.408
S., Central America and Caribbean	0.0058	0.0401	0.000	0.0141	0.0500	0.000
Europe	0.0986	0.2014	0.000	0.0736	0.1809	0.000
Africa	0.0044	0.0099	0.003	0.0118	0.0181	0.062
Asia	0.0543	0.0783	0.000	0.0437	0.1292	0.000
Oceania	0.0041	0.0011	0.004	0.0012	0.0009	0.521

Length of Time in Canada since Immigration						
Non-Immigrant	0.6810	0.6189	0.000	0.8480	0.6120	0.000
20 or more	0.2604	0.3072	0.000	0.1196	0.2993	0.000
10 to 19	0.0317	0.0477	0.001	0.0211	0.0597	0.000
0 to 9	0.0268	0.0262	0.899	0.0113	0.0290	0.001
Highest Level of Education						
Less than Secondary School	0.2489	0.2000	0.000	0.2103	0.1708	0.000
Secondary School	0.1727	0.1818	0.181	0.1591	0.1875	0.001
Some Post-Secondary	0.0616	0.0618	0.966	0.0665	0.0612	0.350
Post-Secondary	0.5168	0.5563	0.000	0.5641	0.5805	0.148
Household Income*						
Quintile 1 (Lowest)	0.1988	0.1597	0.000	0.1794	0.1822	0.775
Quintile 2	0.1995	0.1838	0.029	0.2095	0.1900	0.025
Quintile 3	0.1958	0.1967	0.902	0.2004	0.1900	0.220
Quintile 4	0.1889	0.2013	0.080	0.1964	0.2106	0.106
Quintile 5 (highest)	0.2170	0.2584	0.000	0.2142	0.2273	0.184
Body Mass Index (18+) Self Report						
Normal Weight	0.3718	0.3718	1.000	0.3772	0.3746	0.818
Underweight	0.0099	0.0113	0.442	0.0151	0.0118	0.282
Overweight	0.4095	0.4114	0.831	0.3847	0.3939	0.417
Obese-Class I	0.1573	0.1549	0.709	0.1601	0.1640	0.626
Obese-Class II	0.0366	0.0354	0.706	0.0437	0.0383	0.145
Obese-Class III	0.0150	0.0152	0.919	0.0192	0.0173	0.445
Type of Smoker						
Daily	0.1550	0.1558	0.883	0.1860	0.1578	0.000
Occasional	0.0273	0.0298	0.397	0.0331	0.0285	0.249
Never	0.8177	0.8144	0.603	0.7810	0.8137	0.000
Leisure Physical Activity Index						
Active	0.2354	0.2280	0.322	0.2250	0.2185	0.485
Moderately Active	0.2683	0.2684	0.995	0.2545	0.2463	0.411
Inactive	0.4963	0.5036	0.400	0.5205	0.5352	0.213
Frequency of Heavy Drinking (≥5 drinks at one sitting)						
Never	0.7439	0.7472	0.657	0.7016	0.7424	0.000
Less than Once a Month	0.1413	0.1386	0.654	0.1661	0.1368	0.001
Once a Month	0.0395	0.0378	0.601	0.0422	0.0412	0.846
2-3 Times a Month	0.0261	0.0281	0.440	0.0289	0.0295	0.883
Once a Week	0.0311	0.0259	0.058	0.0360	0.0285	0.055
More than Once a Week	0.0181	0.0224	0.058	0.0251	0.0216	0.237
Self Perceived Health						
Poor	0.0358	0.0369	0.730	0.0325	0.0499	0.007
Fair	0.1168	0.1138	0.593	0.1115	0.1131	0.815
Good	0.3021	0.2978	0.592	0.3299	0.3054	0.024
Very Good	0.3634	0.3585	0.581	0.3374	0.3378	0.975

Excellent	0.1819	0.1930	0.109	0.1888	0.1939	0.575
Number of GP Consultations in Past Year						
0 to 3	0.6227	0.6571	0.000	0.7298	0.6786	0.000
4 to 10	0.3040	0.2887	0.062	0.2298	0.2720	0.000
11 to 19	0.0577	0.0443	0.000	0.0309	0.0401	0.042
20 or more	0.0156	0.0098	0.005	0.0094	0.0094	0.996
Number of Specialist Consultations Past Year						
0 to 3	0.9495	0.9360	0.002	0.9477	0.9248	0.000
4 to 10	0.0395	0.0499	0.004	0.0413	0.0605	0.000
11 to 19	0.0051	0.0081	0.076	0.0067	0.0080	0.418
20 or more	0.0059	0.0059	0.991	0.0044	0.0067	0.121
Self Perceived Mental Health						
Excellent	0.3534	0.3927	0.000	0.3832	0.3811	0.856
Very Good	0.3630	0.3647	0.845	0.3489	0.3562	0.487
Good	0.2374	0.1966	0.000	0.2179	0.2054	0.193
Fair	0.0382	0.0354	0.335	0.0414	0.0431	0.737
Poor	0.0079	0.0105	0.198	0.0087	0.0142	0.066
Self Perceived Stress						
Not at all Stressful	0.1715	0.1469	0.000	0.1715	0.1376	0.000
Not Very Stressful	0.2905	0.2731	0.021	0.2695	0.2604	0.335
A Bit Stressful	0.3697	0.3919	0.009	0.3780	0.3850	0.508
Quite a Bit Stressful	0.1395	0.1570	0.008	0.1544	0.1770	0.015
Extremely Stressful	0.0288	0.0310	0.483	0.0265	0.0401	0.002
Year						
2003	0.6024	0.1527	0.000	0.0000	0.0000	0.000
2005	0.3015	0.5363	0.000	0.0000	0.0000	0.000
2007	0.0961	0.3110	0.000	0.0000	0.0000	0.000
2008	0.0000	0.0000	0.000	0.8593	0.5160	0.000
2009	0.0000	0.0000	0.000	0.1408	0.4840	0.000
Age						
50-64	0.7396	0.7316		0.7492	0.7401	
65-74	0.2604	0.2684	0.159	0.2508	0.2599	0.148
Sex						
Male	0.5304	0.5200	0.136	0.5234	0.5339	0.219
Female	0.4696	0.4800		0.4766	0.4661	
Ethnicity						
Caucasian	0.9151	0.8554	0.000	0.9054	0.7933	0.000
Other	0.0849	0.1446		0.0946	0.2067	
Geography						
Urban	0.6914	0.8278	0.000	0.7348	0.8298	0.000
Rural	0.3086	0.1722		0.2652	0.1702	

Ever had a Flu Shot						
Yes	0.5270	0.7074	0.000	0.5664	0.7149	0.000
No	0.4730	0.2926		0.4336	0.2851	
Has Regular Medical Doctor						
Yes	0.9328	0.9443	0.007	0.8823	0.9470	0.000
No	0.0672	0.0557		0.1177	0.0530	
Screening in Past Year: Total						
Yes	0.1060	0.1779	0.000	0.1364	0.2772	0.000
No	0.8940	0.8221		0.8636	0.7228	
Screening in Past Year: gFOBT						
Yes	0.0836	0.1354	0.000	0.1064	0.2222	0.000
No	0.9164	0.8646		0.8936	0.7778	
Screening in Past Year: Endoscopy						
Yes	0.0328	0.0599	0.000	0.0426	0.0829	0.000
No	0.9672	0.9401		0.9574	0.9171	
Screening Adherence: FOBT in Previous 2 Years or Endoscopy in Previous 5 Years						
Yes	0.2203	0.3381	0.000	0.2777	0.4863	0.000
No	0.7797	0.6619		0.7223	0.5137	
Screening Adherence: gFOBT Previous 2 Years						
Yes	0.1322	0.2059	0.000	0.1705	0.3274	0.000
No	0.8678	0.7941		0.8295	0.6726	
Screening Adherence: Endoscopy Previous 5 Years						
Yes	0.1220	0.1893	0.000	0.1495	0.2567	0.000
No	0.8780	0.8107		0.8505	0.7433	

Using CCHS 2003, 2005, 2007, 2008, 2009
Pre-intervention: N=30484, Weighted N=1468619; Post-intervention: N=24960, Weighted N=1270704
All proportions are weighted
Proportions are based on valid responses (complete case analysis used)
Proportions based on closest approximation of average risk population available ages 50-74
**Household Income distribution based on national composition excluding Canadian territories*

C. Period Comparisons for DD and DDD Analyses

Variable	Ontario			Control Provinces		
	Pre Intervention	**Post Intervention**	**p**	**Pre Intervention**	**Post Intervention**	**p**
Screening in Past Year: Total						
Yes	0.1779	0.2772	0.000	0.1060	0.1364	0.000
No	0.8221	0.7228		0.8940	0.8636	
Screening in Past Year: gFOBT						
Yes	0.1354	0.2222	0.000	0.0836	0.1064	0.000
No	0.8646	0.7778		0.9164	0.8936	
Screening in Past Year: Endoscopy						
Yes	0.0599	0.0829	0.000	0.0328	0.0426	0.003
No	0.9401	0.9171		0.9672	0.9574	
Screening Adherence: FOBT in Previous 2 Years or Endoscopy in Previous 5 Years						
Yes	0.3381	0.4863	0.000	0.2203	0.2777	0.000
No	0.6619	0.5137		0.7797	0.7223	
Screening Adherence: gFOBT Previous 2 Years						
Yes	0.2059	0.3274	0.000	0.1322	0.1705	0.000
No	0.7941	0.6726		0.8678	0.8295	
Screening Adherence: Endoscopy Previous 5 Years						
Yes	0.1893	0.2567	0.000	0.1220	0.1495	0.000
No	0.8107	0.7433		0.8780	0.8505	

Using CCHS 2003, 2005, 2007, 2008, 2009
Ontario: N=29489, Weighted N=1683381; Control Provinces: N=25955, Weighted N=1055941
All proportions are weighted
Proportions are based on valid responses (complete case analysis used)
Proportions based on closest approximation of average risk population available ages 50-74
**Household Income distribution based on national composition excluding Canadian territories*

D. Predictors of Past-Year Screening: Multivariate Results by Intervention Period and Group

Variable	Ontario 2003-7				Control Provinces 2003-7				Ontario 2008-9				Control Provinces 2008-9			
		95% CI				95% CI				95% CI				95% CI		
	Mrg Effect	Lower	Upper	p	Mrg Effect	Lower	Upper	p	Mrg Effect	Lower	Upper	p	Mrg Effect	Lower	Upper	p
Age																
50-64	ref				ref				ref				ref			
65-74	0.0267	0.0067	0.0466	0.009	0.0494	0.0294	0.0694	0.000	0.0599	0.0298	0.0900	0.000	0.0417	0.0203	0.0630	0.000
Sex																
Male	0.0166	-0.0024	0.0356	0.087	0.0249	0.0084	0.0414	0.003	-0.0057	-0.0354	0.0240	0.706	0.0198	0.0018	0.0378	0.031
Female	ref				ref				ref				ref			
Marital Status																
Married	ref				ref				ref				ref			
Common-Law	-0.0314	-0.0661	0.0032	0.075	0.0123	-0.0294	0.0540	0.564	-0.0053	-0.0651	0.0545	0.862	0.0115	-0.0296	0.0525	0.584
Widowed	-0.0226	-0.0523	0.0070	0.134	0.0023	-0.0280	0.0326	0.882	-0.0385	-0.0928	0.0159	0.165	-0.0190	-0.0471	0.0090	0.184
Separated	-0.0698	-0.1090	-0.0306	0.000	-0.0221	-0.0546	0.0104	0.182	0.0271	-0.0485	0.1027	0.482	-0.0169	-0.0577	0.0239	0.418
Divorced	-0.0285	-0.0588	0.0018	0.066	-0.0023	-0.0267	0.0221	0.853	-0.0326	-0.0826	0.0174	0.201	0.0123	-0.0218	0.0465	0.480
Single, Never Married	-0.0345	-0.0629	-0.0061	0.017	-0.0199	-0.0463	0.0065	0.140	0.0146	-0.0343	0.0635	0.559	0.0077	-0.0259	0.0414	0.653
Ethnicity																
Caucasian	0.0035	-0.0383	0.0452	0.871	0.0491	0.0257	0.0725	0.000	0.0102	-0.0533	0.0737	0.752	0.0328	0.0012	0.0645	0.042
Other	ref				ref				ref				ref			
Highest Level of Education																
Less than Secondary School	ref				ref				ref				ref			
Secondary School	0.0101	-0.0152	0.0353	0.435	0.0075	-0.0164	0.0315	0.537	-0.0015	-0.0483	0.0452	0.949	-0.0042	-0.0319	0.0234	0.764
Some Post-Secondary	0.0617	0.0162	0.1071	0.008	0.0102	-0.0260	0.0463	0.581	0.0563	-0.0110	0.1236	0.101	-0.0222	-0.0536	0.0092	0.166
Post-Secondary	0.0275	0.0054	0.0496	0.015	-0.0028	-0.0207	0.0152	0.762	0.0124	-0.0264	0.0512	0.531	0.0073	-0.0146	0.0292	0.513
Household Income																
Quintile 1 (Lowest)	ref				ref				ref				ref			

Quintile 2	0.0179	-0.0114	0.0472	0.232	-0.0001	-0.0193	0.0191	0.988	0.0206	-0.0284	0.0696	0.411	0.0007	-0.0290	0.0304	0.964
Quintile 3	0.0312	0.0020	0.0604	0.036	0.0164	-0.0065	0.0392	0.160	0.0236	-0.0237	0.0710	0.327	-0.0007	-0.0296	0.0283	0.964
Quintile 4	0.0220	-0.0079	0.0518	0.150	0.0219	-0.0021	0.0460	0.074	0.0409	-0.0084	0.0902	0.104	-0.0008	-0.0315	0.0299	0.959
Quintile 5 (highest)	0.0129	-0.0175	0.0434	0.405	0.0257	0.0004	0.0510	0.047	0.0500	0.0002	0.0997	0.049	0.0075	-0.0260	0.0410	0.662
Geography																
Urban	0.0014	-0.0162	0.0190	0.874	-0.0208	-0.0368	-0.0048	0.011	0.0092	-0.0187	0.0371	0.519	-0.0098	-0.0278	0.0081	0.284
Rural	ref				ref				ref				ref			
Body Mass Index (18+) Self Report																
Normal Weight	ref				ref				ref				ref			
Underweight	-0.0307	-0.0976	0.0363	0.369	-0.0039	-0.0742	0.0664	0.913	0.0724	-0.0846	0.2294	0.366	-0.0724	-0.1205	-0.0244	0.003
Overweight	-0.0182	-0.0383	0.0018	0.075	-0.0063	-0.0238	0.0112	0.480	-0.0028	-0.0355	0.0298	0.864	0.0016	-0.0194	0.0226	0.882
Obese-Class I	-0.0255	-0.0514	0.0004	0.053	-0.0027	-0.0249	0.0195	0.810	-0.0202	-0.0615	0.0211	0.338	-0.0215	-0.0460	0.0030	0.086
Obese-Class II	-0.0069	-0.0454	0.0317	0.727	-0.0419	-0.0687	-0.0151	0.002	0.0571	-0.0052	0.1193	0.073	-0.0395	-0.0740	-0.0049	0.025
Obese-Class III	-0.0377	-0.1084	0.0331	0.296	-0.0536	-0.0941	-0.0131	0.009	0.0110	-0.0841	0.1062	0.820	-0.0416	-0.0960	0.0129	0.135
Type of Smoker																
Daily	ref				ref				ref				ref			
Occasional	0.0269	-0.0389	0.0927	0.423	0.0158	-0.0283	0.0600	0.482	-0.0382	-0.1259	0.0494	0.392	-0.0142	-0.0592	0.0309	0.538
Never	0.0347	0.0122	0.0573	0.003	0.0349	0.0170	0.0529	0.000	0.0263	-0.0147	0.0673	0.208	0.0104	-0.0124	0.0332	0.371
Leisure Physical Activity Index																
Active	ref				ref				ref				ref			
Moderately Active	-0.0022	-0.0280	0.0236	0.866	-0.0133	-0.0367	0.0101	0.266	-0.0177	-0.0559	0.0205	0.364	-0.0027	-0.0284	0.0230	0.837
Inactive	-0.0462	-0.0693	-0.0230	0.000	-0.0341	-0.0539	-0.0142	0.001	-0.0767	-0.1102	-0.0432	0.000	-0.0323	-0.0546	-0.0100	0.004
Frequency of Heavy Drinking (≥5 drinks at one sitting)																
Never	ref				ref				ref				ref			
Less than Once a Month	-0.0008	-0.0269	0.0252	0.951	0.0045	-0.0186	0.0275	0.704	-0.0353	-0.0746	0.0039	0.078	0.0012	-0.0226	0.0250	0.921
Once a Month	0.0227	-0.0214	0.0667	0.313	0.0107	-0.0344	0.0559	0.641	-0.0281	-0.1160	0.0598	0.531	-0.0146	-0.0517	0.0225	0.440
2-3 Times a Month	-0.0334	-0.0727	0.0058	0.095	0.0217	-0.0299	0.0733	0.410	-0.0619	-0.1372	0.0134	0.107	-0.0233	-0.0652	0.0186	0.277
Once a Week	0.0121	-0.0415	0.0657	0.659	0.0312	-0.0300	0.0924	0.318	0.0078	-0.0768	0.0923	0.857	-0.0266	-0.0718	0.0185	0.248

More than Once a Week	0.0147	-0.0470	0.0764	0.641	-0.0141	-0.0657	0.0375	0.592	0.0991	-0.0001	0.1983	0.050	-0.0210	-0.0701	0.0282	0.403
Self Perceived Health																
Poor	ref				ref				ref				ref			
Fair	0.0174	-0.0315	0.0664	0.486	0.0232	-0.0097	0.0562	0.167	0.0095	-0.0818	0.1008	0.838	0.0077	-0.0466	0.0620	0.781
Good	0.0026	-0.0429	0.0482	0.909	0.0157	-0.0155	0.0469	0.323	-0.0063	-0.0960	0.0835	0.891	-0.0258	-0.0768	0.0252	0.321
Very Good	-0.0064	-0.0535	0.0406	0.789	0.0106	-0.0217	0.0429	0.521	-0.0096	-0.1009	0.0817	0.837	-0.0287	-0.0814	0.0239	0.285
Excellent	0.0062	-0.0450	0.0575	0.811	0.0477	0.0093	0.0861	0.015	0.0018	-0.0952	0.0987	0.971	-0.0104	-0.0669	0.0462	0.719
Ever had a Flu Shot																
Yes	0.0650	0.0431	0.0869	0.000	0.0325	0.0163	0.0487	0.000	0.0945	0.0615	0.1276	0.000	0.0469	0.0277	0.0662	0.000
No	ref				ref				ref				ref			
Has Regular Medical Doctor																
Yes	0.1007	0.0438	0.1576	0.001	0.0685	0.0237	0.1133	0.003	0.2820	0.2138	0.3502	0.000	0.1249	0.0752	0.1747	0.000
No	ref				ref				ref				ref			
Number of GP Consultations in Past Year																
0 to 3	ref				ref				ref				ref			
4 to 10	0.0378	0.0169	0.0587	0.000	0.0433	0.0251	0.0616	0.000	0.0848	0.0526	0.1170	0.000	0.0398	0.0184	0.0611	0.000
11 to 19	0.0814	0.0338	0.1290	0.001	0.0699	0.0329	0.1069	0.000	0.0889	0.0147	0.1631	0.019	0.0303	-0.0134	0.0740	0.175
20 or more	0.0714	-0.0126	0.1554	0.096	0.0596	-0.0006	0.1199	0.052	0.0843	-0.0532	0.2217	0.229	0.0498	-0.0332	0.1327	0.240
Number of Specialist Consultations Past Year																
0 to 3	ref				ref				ref				ref			
4 to 10	0.0360	-0.0057	0.0777	0.090	-0.0051	-0.0396	0.0293	0.770	0.0205	-0.0418	0.0828	0.519	0.0360	-0.0118	0.0838	0.140
11 to 19	-0.1152	-0.1569	-0.0735	0.000	0.0069	-0.0843	0.0982	0.882	-0.0079	-0.1373	0.1214	0.904	-0.0063	-0.0985	0.0858	0.893
20 or more	0.0314	-0.1220	0.1849	0.688	0.0561	-0.0551	0.1674	0.323	-0.0646	-0.2179	0.0887	0.409	0.0808	-0.0465	0.2081	0.214
Self Perceived Mental Health																
Excellent	ref				ref				ref				ref			
Very Good	-0.0107	-0.0311	0.0098	0.306	-0.0033	-0.0218	0.0152	0.726	-0.0002	-0.0335	0.0331	0.990	0.0036	-0.0178	0.0249	0.744
Good	-0.0195	-0.0447	0.0057	0.129	-0.0050	-0.0257	0.0157	0.637	0.0155	-0.0252	0.0561	0.456	-0.0208	-0.0443	0.0028	0.084

Fair	-0.0182	-0.0630	0.0265	0.424	0.0101	-0.0372	0.0573	0.677	-0.0344	-0.1111	0.0424	0.380	-0.0223	-0.0707	0.0261	0.366
Poor	-0.0777	-0.1597	0.0042	0.063	-0.0002	-0.0778	0.0773	0.995	-0.0524	-0.1851	0.0802	0.438	0.0300	-0.1098	0.1698	0.674
Self Perceived Stress																
Not at all Stressful	ref				ref				ref				ref			
Not Very Stressful	0.0183	-0.0068	0.0434	0.152	-0.0077	-0.0296	0.0141	0.487	0.0149	-0.0260	0.0558	0.475	-0.0099	-0.0349	0.0150	0.436
A Bit Stressful	0.0072	-0.0171	0.0315	0.562	0.0053	-0.0177	0.0283	0.653	0.0224	-0.0183	0.0632	0.281	-0.0091	-0.0349	0.0168	0.492
Quite a Bit Stressful	0.0118	-0.0199	0.0436	0.465	-0.0189	-0.0463	0.0085	0.177	-0.0025	-0.0537	0.0488	0.924	-0.0194	-0.0506	0.0118	0.223
Extremely Stressful	0.0807	0.0160	0.1455	0.015	0.0652	-0.0242	0.1546	0.153	-0.0624	-0.1445	0.0197	0.136	0.0135	-0.0429	0.0699	0.639

Using CCHS 2003, 2005, 2007, 2008, 2009
Dependent variable is self-report of having a gFOBT or endoscopy (FS or colonoscopy) in past year
Ontario 2003-7: N=18647, Weighted N=1039602
Control Provinces 2003-7: N=11837, Weighted N=429016
Ontario 2008-9: N=10842, Weighted N=643778
Control Provinces 2008-9: N=14118, Weighted N=626924
All proportions are weighted
ref=reference standard
**Household Income distribution based on national composition excluding Canadian territories*
Results are controlled for all other covariates in model. Results shown are average marginal effects calculated from multivariate logistic regression model
Marginal Effects are based on valid responses (complete case analysis used)
Marginal Effects based on closest approximation of average risk population available ages 50-74

E. Difference in Differences Analysis

Variable	Outcome: gFOBT Past Year				Outcome: Endoscopy Past Year			
		95% CI				95% CI		
	Mrg Effect	Lower	Upper	p	Mrg Effect	Lower	Upper	p
Intervention Group	0.0638	0.0412	0.0864	0.000	0.0089	-0.0042	0.0219	0.182
Post-Intervention Period	0.0744	0.0535	0.0953	0.000	0.0234	0.0087	0.0381	0.002
Group*Period	0.0516	0.0319	0.0712	0.000	0.0088	-0.0049	0.0226	0.207
Age								
50-64	ref				ref			
65-74	0.0336	0.0232	0.0440	0.000	0.0164	0.0092	0.0237	0.000
Sex								
Male	0.0105	0.0011	0.0199	0.028	0.0087	0.0022	0.0151	0.009
Female	ref				ref			
Ethnicity								
Caucasian	0.0074	-0.0095	0.0242	0.394	0.0103	0.0009	0.0198	0.031
Other	ref				ref			
Highest Level of Education								
Less than Secondary School	ref				ref			
Secondary School	0.0044	-0.0100	0.0189	0.548	-0.0003	-0.0096	0.0091	0.957
Some Post-Secondary	0.0134	-0.0079	0.0348	0.217	0.0162	-0.0033	0.0358	0.104
Post-Secondary	0.0082	-0.0038	0.0202	0.179	0.0075	-0.0001	0.0151	0.052
Household Income*								
Quintile 1 (Lowest)	ref				ref			
Quintile 2	0.0132	-0.0024	0.0288	0.097	-0.0020	-0.0107	0.0067	0.655
Quintile 3	0.0188	0.0035	0.0342	0.016	0.0082	-0.0016	0.0180	0.102
Quintile 4	0.0090	-0.0064	0.0244	0.251	0.0168	0.0059	0.0276	0.002
Quintile 5 (highest)	0.0143	-0.0017	0.0304	0.079	0.0193	0.0080	0.0305	0.001
Geography								
Urban	-0.0050	-0.0142	0.0042	0.286	0.0022	-0.0039	0.0084	0.472
Rural	ref				ref			
Type of Smoker								
Daily	ref				ref			
Occasional	-0.0042	-0.0316	0.0231	0.761	0.0122	-0.0130	0.0375	0.342
Never	0.0213	0.0091	0.0335	0.001	0.0092	0.0008	0.0177	0.032
Leisure Physical Activity Index								
Active	ref				ref			
Moderately Active	-0.0094	-0.0233	0.0044	0.182	-0.0063	-0.0167	0.0041	0.234
Inactive	-0.0439	-0.0560	-0.0318	0.000	-0.0178	-0.0262	-0.0094	0.000
Self Perceived Health								
Poor	ref				ref			

Fair	0.0033	-0.0242	0.0309	0.812	0.0052	-0.0119	0.0223	0.551
Good	-0.0034	-0.0294	0.0227	0.801	-0.0094	-0.0253	0.0065	0.246
Very Good	-0.0027	-0.0292	0.0237	0.839	-0.0104	-0.0265	0.0058	0.210
Excellent	0.0115	-0.0170	0.0401	0.429	-0.0052	-0.0237	0.0133	0.582
Ever had a Flu Shot								
Yes	0.0491	0.0380	0.0603	0.000	0.0213	0.0135	0.0292	0.000
No	ref				ref			
Has Regular Medical Doctor								
Yes	0.1249	0.0976	0.1521	0.000	0.0446	0.0216	0.0676	0.000
No	ref				ref			
Number of GP Consultations in Past Year								
0 to 3	ref				ref			
4 to 10	0.0367	0.0252	0.0481	0.000	0.0212	0.0133	0.0292	0.000
11 to 19	0.0516	0.0277	0.0756	0.000	0.0446	0.0275	0.0617	0.000
20 or more	0.0480	0.0127	0.0834	0.008	0.0649	0.0329	0.0969	0.000

Using CCHS 2003, 2005, 2007, 2008, 2009
N=58142, Weighted N= 2882630; All proportions are weighted
Dependent variable is self-report of having a gFOBT in past year (left) and endoscopy (flexible sigmoidoscopy or colonoscopy) in past year (right)
ref=reference standard
**Household Income distribution based on national composition excluding Canadian territories*
Covariates for province and year of CCHS also included in model. Are estimable in logistic regression but not estimable when translating results into marginal effects and are omitted here
Results are controlled for all other covariates in model. Results shown are average marginal effects calculated from multivariate logistic regression model
Proportions are based on valid responses (complete case analysis used)
Proportions based on closest approximation of average risk population available ages 50-74

F. Difference in Difference in Differences Analysis

Outcome		DDD*			
		Mrg Effect	95% CI		p
gFOBT	*T*	0.0024	-0.0780	0.0829	0.952
	G	0.0578	-0.0273	0.1429	0.183
	MD	0.0801	0.0151	0.1450	0.016
	*G*T*	0.0434	-0.0587	0.1454	0.405
	*T*MD*	0.0739	-0.0067	0.1546	0.072
	*G*MD*	0.0065	-0.0740	0.0870	0.874
	*T*G*MD*	0.0078	-0.0953	0.1108	0.883
Endoscopy	*T*	0.0453	-0.0051	0.0956	0.078
	G	0.0216	-0.0280	0.0712	0.393
	MD	0.0512	0.0201	0.0822	0.001
	*G*T*	-0.0548	-0.1252	0.0156	0.127
	*T*MD*	-0.0227	-0.0739	0.0284	0.384
	*G*MD*	-0.0130	-0.0628	0.0368	0.609
	*T*G*MD*	0.0651	-0.0069	0.1372	0.077

Using CCHS 2003, 2005, 2007, 2008, 2009
N=58142, Weighted N= 2 882 630. All results are weighted.
Dependent variable is self-report of having a gFOBT in past year or endoscopy (flexible sigmoidoscopy or colonoscopy) in past year. G is a group indicator=1 if province is Ontario and 0 otherwise, T is a time indicator =1 if year≥2008 and 0 otherwise. Modifying variables include MD for reporting having a regular medical doctor.
**: main effects controlled for year, province, sex, age category, geography, self-rated health, reporting flu shot, physical activity index, smoking status, ethnicity, education, income, #GP consultations past year.*
Results shown are average marginal effects calculated from multivariate logistic regression model
Marginal Effects are based on valid responses (complete case analysis used)
Marginal Effects based on closest approximation of average risk population available ages 50-74

Appendix III

A. Comparison of Sample Characteristics in RDD Models

Variable	Bandwidth 40-60 Years			Bandwidth 45-55Years		
	$<Th_i$	$>=Th_i$	p	$<Th_i$	$>=Th_i$	p
Sex						
Male	0.5136	0.5247	0.102	0.5033	0.5245	0.024
Female	0.4864	0.4753		0.4967	0.4755	
Marital Status						
Married	0.6885	0.7194	0.000	0.6822	0.7104	0.001
Common-Law	0.0894	0.0656	0.000	0.0863	0.0716	0.003
Widowed	0.0070	0.0260	0.000	0.0105	0.0181	0.002
Separated	0.0436	0.0330	0.000	0.0449	0.0366	0.047
Divorced	0.0624	0.0842	0.000	0.0708	0.0853	0.004
Single, Never Married	0.1091	0.0719	0.000	0.1054	0.0780	0.000
Country of Birth						
Canada	0.7332	0.7256	0.307	0.7448	0.7352	0.364
Other North America	0.0130	0.0154	0.127	0.0150	0.0139	0.673
S., Central America and Caribbean	0.0355	0.0352	0.941	0.0388	0.0365	0.671
Europe	0.0865	0.1208	0.000	0.0877	0.1032	0.021
Africa	0.0136	0.0124	0.552	0.0114	0.0132	0.519
Asia	0.1164	0.0893	0.000	0.1012	0.0970	0.610
Oceania	0.0019	0.0013	0.243	0.0010	0.0010	0.971
Ethnicity						
Caucasian	0.7965	0.8384	0.000	0.8117	0.8258	0.137
Other	0.2035	0.1616		0.1883	0.1742	
Highest Level of Education						
Less than Secondary School	0.0968	0.1409	0.000	0.1091	0.1262	0.005
Secondary School	0.1718	0.1869	0.003	0.1829	0.1928	0.199
Some Post-Secondary	0.0648	0.0656	0.815	0.0648	0.0658	0.843
Post-Secondary	0.6666	0.6066	0.000	0.6431	0.6151	0.005
Household Income						
Quintile 1 (Lowest)	0.1496	0.1352	0.005	0.1465	0.1261	0.009
Quintile 2	0.1733	0.1556	0.001	0.1623	0.1544	0.327
Quintile 3	0.2034	0.1897	0.012	0.1952	0.1885	0.395
Quintile 4	0.2281	0.2230	0.371	0.2334	0.2209	0.119
Quintile 5 (highest)	0.2457	0.2965	0.000	0.2626	0.3100	0.000
Geography						
Urban	0.8086	0.7934	0.001	0.8004	0.7970	0.614
Rural	0.1914	0.2066		0.1996	0.2030	

Body Mass Index (18+) Self Report						
Normal Weight	0.4364	0.3799	0.000	0.4265	0.3936	0.001
Underweight	0.0154	0.0111	0.014	0.0154	0.0120	0.212
Overweight	0.3691	0.3940	0.000	0.3762	0.3837	0.445
Obese-Class I	0.1288	0.1576	0.000	0.1305	0.1528	0.001
Obese-Class II	0.0353	0.0386	0.154	0.0366	0.0376	0.752
Obese-Class III	0.0150	0.0189	0.012	0.0148	0.0202	0.012
Type of Smoker						
Daily	0.2132	0.1879	0.000	0.2142	0.2007	0.098
Occasional	0.0479	0.0340	0.000	0.0396	0.0349	0.220
Never	0.7389	0.7781	0.000	0.7462	0.7645	0.043
Leisure Physical Activity Index						
Active	0.2384	0.2224	0.004	0.2385	0.2228	0.049
Moderately Active	0.2597	0.2551	0.476	0.2657	0.2543	0.223
Inactive	0.5020	0.5225	0.005	0.4958	0.5229	0.009
Frequency of Heavy Drinking (≥5 drinks at one sitting)						
Never	0.5961	0.6852	0.000	0.6042	0.6568	0.000
Less than Once a Month	0.2209	0.1715	0.000	0.2076	0.1874	0.011
Once a Month	0.0592	0.0476	0.000	0.0629	0.0516	0.015
2-3 Times a Month	0.0484	0.0344	0.000	0.0487	0.0350	0.001
Once a Week	0.0460	0.0357	0.000	0.0465	0.0405	0.118
More than Once a Week	0.0294	0.0256	0.059	0.0300	0.0287	0.665
Self Perceived Health						
Poor	0.0191	0.0397	0.000	0.0222	0.0355	0.001
Fair	0.0692	0.1013	0.000	0.0754	0.0912	0.004
Good	0.2756	0.2936	0.004	0.2771	0.3003	0.009
Very Good	0.4026	0.3628	0.000	0.3882	0.3715	0.092
Excellent	0.2335	0.2026	0.000	0.2371	0.2014	0.000
Ever had a Flu Shot						
Yes	0.4934	0.5832	0.000	0.5028	0.5615	0.000
No	0.5066	0.4168		0.4972	0.4385	
Has Regular Medical Doctor						
Yes	0.8908	0.9166	0.000	0.8927	0.9132	0.001
No	0.1092	0.0834		0.1073	0.0868	
Number of GP Consultations in Past Year						
0 to 3	0.7690	0.7024	0.000	0.7606	0.7139	0.000
4 to 10	0.1857	0.2437	0.000	0.1923	0.2341	0.000
11 to 19	0.0298	0.0397	0.000	0.0338	0.0373	0.300
20 or more	0.0155	0.0142	0.434	0.0134	0.0147	0.589
Number of Specialist Consultations Past Year						
0 to 3	0.9494	0.9416	0.022	0.9493	0.9424	0.159

4 to 10	0.0361	0.0445	0.005	0.0354	0.0436	0.057
11 to 19	0.0084	0.0073	0.413	0.0089	0.0069	0.317
20 or more	0.0061	0.0066	0.643	0.0064	0.0071	0.698
Self Perceived Mental Health						
Excellent	0.3777	0.3749	0.688	0.3805	0.3704	0.317
Very Good	0.3611	0.3619	0.900	0.3535	0.3654	0.195
Good	0.2030	0.2060	0.607	0.2061	0.2066	0.954
Fair	0.0466	0.0423	0.134	0.0483	0.0421	0.135
Poor	0.0116	0.0150	0.049	0.0115	0.0155	0.102
Self Perceived Stress						
Not at all Stressful	0.0670	0.1050	0.000	0.0718	0.0918	0.000
Not Very Stressful	0.1888	0.2299	0.000	0.1870	0.2044	0.026
A Bit Stressful	0.4617	0.4203	0.000	0.4578	0.4409	0.091
Quite a Bit Stressful	0.2393	0.2034	0.000	0.2387	0.2202	0.031
Extremely Stressful	0.0433	0.0414	0.526	0.0448	0.0426	0.606
Province of Residence						
NFLD	0.0337	0.0374	0.017	0.0349	0.0359	0.652
PEI	0.0057	0.0064	0.112	0.0058	0.0064	0.347
NS	0.0341	0.0372	0.099	0.0344	0.0359	0.601
NB	0.0275	0.0297	0.139	0.0275	0.0279	0.854
QC	0.0659	0.0816	0.000	0.0722	0.0832	0.076
ON	0.6468	0.6143	0.000	0.6332	0.6118	0.014
MAN	0.0094	0.0097	0.830	0.0110	0.0099	0.537
SASK	0.0298	0.0355	0.000	0.0321	0.0365	0.063
ALTA	0.0284	0.0323	0.094	0.0319	0.0340	0.584
BC	0.1187	0.1159	0.486	0.1170	0.1184	0.804
Year						
2003	0.1666	0.1520	0.000	0.1614	0.1571	0.480
2005	0.2883	0.2555	0.000	0.2776	0.2535	0.006
2007	0.1259	0.1351	0.032	0.1272	0.1377	0.101
2008	0.2882	0.3178	0.000	0.2985	0.3106	0.189
2009	0.1309	0.1396	0.063	0.1353	0.1411	0.407
Screening in Past Year: Total						
Yes	0.0456	0.1525	0.000	0.0574	0.1355	0.000
No	0.9544	0.8475		0.9426	0.8645	
Screening in Past Year: gFOBT						
Yes	0.0330	0.1187	0.000	0.0447	0.1067	0.000
No	0.9670	0.8813		0.9553	0.8933	
Screening in Past Year: Endoscopy						
Yes	0.0174	0.0480	0.000	0.0192	0.0416	0.000
No	0.9826	0.9520		0.9808	0.9584	

Screening Adherence:						
FOBT in Previous 2 Years or Endoscopy in Previous 5 Years						
Yes	0.1034	0.2899	0.000	0.1247	0.2493	0.000
No	0.8966	0.7101		0.8753	0.7507	
Screening Adherence:						
gFOBT Previous 2 Years						
Yes	0.0540	0.1792	0.000	0.0693	0.1554	0.000
No	0.9460	0.8208		0.9307	0.8446	
Screening Adherence:						
Endoscopy Previous 5 Years						
Yes	0.0628	0.1541	0.000	0.0727	0.1291	0.000
No	0.9372	0.8459		0.9273	0.8709	

Using CCHS 2003, 2005, 2007, 2008, 2009
Bandwidth 40-60: N= 55997, Weighted N= 3725238
Bandwidth 45-55: N= 28505, Weighted N= 1967749.
All results are weighted.
Dependent variable is self-report of having a gFOBT in past year or endoscopy (flexible sigmoidoscopy or colonoscopy) in past year combined measure.
Bandwidth Indicates range of age in years of participants included in model
Th_i indicates the threshold or cutoff division which equals 1 if participants are aged 50 and older and 0 otherwise. See Equation 5.
Proportions are based on valid responses (complete case analysis used)
Proportions based on closest approximation of average risk population available

B. Proportion Screened by Age by Group and Period with Prediction

	Control Group		Ontario			
Age	**2003-7**	**2008-9**	**2003-7**	**2008-9**	**Pred 1**	**Pred 2**
40	0.031	0.036	0.030	*0.034*		
41	0.046	0.027	0.039	*0.033*		
42	0.037	0.029	0.034	*0.036*		
43	0.042	0.020	0.044	*0.044*		
44	0.050	0.025	0.026	*0.037*		
45	0.034	0.052	0.060	*0.041*		
46	0.037	0.028	0.039	*0.064*		
47	0.024	0.068	0.070	*0.085*	0.063	0.058
48	0.042	0.073	0.082	*0.048*	0.070	0.069
49	0.045	0.052	0.070	*0.090*		
50	0.040	0.050	0.110	*0.142*	0.190	0.176
51	0.077	0.115	0.183	*0.236*		
52	0.058	0.085	0.108	*0.238*	0.209	0.212
53	0.092	0.089	0.136	*0.181*	0.218	0.233
54	0.068	0.105	0.137	*0.269*		
55	0.081	0.116	0.183	*0.266*		
56	0.107	0.129	0.168	*0.266*		
57	0.112	0.111	0.173	*0.223*		
58	0.086	0.127	0.224	*0.249*		
59	0.130	0.161	0.173	*0.293*		
60	0.101	0.171	0.133	*0.290*		
Total	**0.0594**	**0.0754**	**0.0946**	0.1362		

Using CCHS 2003, 2005, 2007, 2008, 2009
N=56055, Weighted N= 3727816. All results are weighted.
Dependent variable is self-report of having a gFOBT in past year or endoscopy (flexible sigmoidoscopy or colonoscopy) in past year combined measure.
Pred 1=outcome predicted with model for selected values of age for Ontario Post-Intervention using linear age terms. Pred 2=outcomes predicted with model for selected ages using squared polynomial age terms.
Proportions are based on valid responses (complete case analysis used)
Proportions based on closest approximation of average risk population available