CARBOHYDRATES AND COLORECTAL CANCER RISK AMONG CHINESE IN NORTH AMERICA

by

MARILYN JEAN BORUGIAN

B.A., University of Michigan, 1970

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

Department of Health Care and Epidemiology

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 2000

© Marilyn Jean Borugian, 2000
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Health Care & Epidemiology

The University of British Columbia
Vancouver, Canada

Date Aug 14, 2002

DE-6 (2/88)
ABSTRACT

CARBOHYDRATES AND
COLORECTAL CANCER RISK
AMONG CHINESE IN NORTH
AMERICA

by Marilyn Jean Borugian

University of British Columbia

Department of Health Care and Epidemiology

Chairperson of the Supervisory Committee: Professor Samuel B. Sheps

Background. Previous studies have analysed total carbohydrates as a dietary risk factor for colorectal cancer (CRC), but obtained conflicting results, perhaps due in part to the embedded negative confounder, fiber. The aim of this study is to analyse the non-fiber ("effective") carbohydrate component (eCarb) separately, and to test the hypothesis that effective carbohydrate consumption is directly related to colorectal cancer risk. Method. The data (473 cases, 1192 controls) are from a large, multi-centre case-control study of migrant Chinese. Multivariate logistic regression was used to perform a secondary analysis, controlling for age, fat, protein, fiber, physical inactivity, calcium, retinol, body mass index, family history, education, years in North America, and for women, parity, age at
menopause, hormone replacement therapy use and oral contraceptive use. The model was used to calculate odds ratios (OR) and 95% confidence intervals (CI) to estimate risk from increasing eCarb consumption among subgroups by gender and cancer site. Results. A statistically significant positive association was observed between eCarb consumption and risk of CRC in both males (OR=2.02 from lowest to highest tertile of eCarb consumption, 95%CI=1.28, 3.19) and females (OR=2.54, 95%CI=1.44, 4.46). A striking gender difference in cancer site affected was observed, with risk concentrated in the right colon for women (OR=7.35, 95%CI=2.60, 20.80), and in the rectum for men (OR=2.78, 95%CI=1.40, 5.54). The other sites for each gender showed little or no significant association of eCarb consumption with risk. Onset of menopause had a significant protective effect, and there was a near-significant interaction of menopause with eCarb consumption (p=0.07). Conclusion. Increased eCarb consumption is associated with increased risk of colorectal cancer in both genders. The right colon risk is greatest in females, while the rectal risk is greatest in males. In females, colorectal cancer risk may be reduced by onset of menopause.
TABLE OF CONTENTS

Abstract ............................................................................................................................................................ ii
List of Tables ................................................................................................................................................ vi
List of Figures .............................................................................................................................................. vii
Acknowledgments ....................................................................................................................................... viii
Glossary ........................................................................................................................................................ ix
Quotation ....................................................................................................................................................... x

INTRODUCTION ........................................................................................................................................... 1

  Purpose ....................................................................................................................................................... 1
  Research Objectives ................................................................................................................................. 2
  Rationale ..................................................................................................................................................... 3
  The Following Sections .............................................................................................................................. 4

BACKGROUND ............................................................................................................................................. 5

  Overview ..................................................................................................................................................... 5
  Diet and Cancer ......................................................................................................................................... 6
    Paleolithic Nutrition .................................................................................................................................. 6
    The Post-Industrial Diet .......................................................................................................................... 10
    The Current Patterns of Diet and Cancer in China and North America .......................................... 12
    Suspected Dietary Risk Factors ............................................................................................................... 14
  Colorectal Cancer .................................................................................................................................... 16
    Description of Sites .................................................................................................................................. 16
    Epidemiology .......................................................................................................................................... 17
    CRC Risk Factors ................................................................................................................................... 18
  The Role of Carbohydrates ....................................................................................................................... 19
    Definition of Carbohydrate Subgroups ................................................................................................. 19
    Glycemic Index ....................................................................................................................................... 21
    Carbohydrates and Blood Insulin Levels ............................................................................................... 23
  Insulin, Insulin Resistance and Hyperinsulinemia ................................................................................... 24
    The Pathway from Carbohydrates to Hyperinsulinemia ......................................................................... 24
    Effects of Excess Insulin ......................................................................................................................... 26
    Insulin and Colorectal Cancer ................................................................................................................. 27
    Insulin and Other Hormones .................................................................................................................. 27
  Reproductive Factors ............................................................................................................................... 28
    Reproductive Risk Factors for CRC ......................................................................................................... 28
    Hormonal Interactions with the Colon .................................................................................................... 29
    The Reversal at Menopause .................................................................................................................... 30
  Summary ...................................................................................................................................................... 31

METHODS ...................................................................................................................................................... 32

  Study Design ............................................................................................................................................ 32
  Cases .......................................................................................................................................................... 33
  Controls ..................................................................................................................................................... 33
  Interviews .................................................................................................................................................. 34
  Diet History ............................................................................................................................................... 34


## LIST OF TABLES

<table>
<thead>
<tr>
<th>Number</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Selected characteristics of colorectal cancer cases and controls by gender.</td>
<td>51</td>
</tr>
<tr>
<td>2. By gender: Association between eCarb intake and colorectal cancer risk.</td>
<td>53</td>
</tr>
<tr>
<td>3. By site and gender: Association between eCarb intake and colorectal cancer risk.</td>
<td>54</td>
</tr>
<tr>
<td>4. By age group: Association between eCarb intake and colorectal cancer risk.</td>
<td>57</td>
</tr>
<tr>
<td>5. By years in NA: Association between eCarb intake and colorectal cancer risk.</td>
<td>58</td>
</tr>
<tr>
<td>6. Female colorectal cancer risk by tertile of eCarb intake including reproductive variables.</td>
<td>59</td>
</tr>
<tr>
<td>7. By menopausal status: Association between eCarb intake and colorectal cancer risk.</td>
<td>61</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Number</th>
<th>Page</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>8</td>
<td>Hunter-gatherer compared with modern Western dietary composition.</td>
</tr>
<tr>
<td>2.</td>
<td>20</td>
<td>Carbohydrate sub-components.</td>
</tr>
<tr>
<td>3.</td>
<td>52</td>
<td>Mean eCarb intake by gender and age group.</td>
</tr>
<tr>
<td>4.</td>
<td>55</td>
<td>CRC risk by tertile of eCarb intake and gender.</td>
</tr>
<tr>
<td>5.</td>
<td>56</td>
<td>Risk by tertile of eCarb intake for rectal, left colon and right colon cancer.</td>
</tr>
<tr>
<td>6.</td>
<td>60</td>
<td>CRC risk for 3rd tertile of eCarb intake by years in North America.</td>
</tr>
<tr>
<td>7.</td>
<td>74</td>
<td>Causal pathway.</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

The author wishes to acknowledge the invaluable help and comments of Dr. Sam Sheps, the generous support and assistance of Mr. Rick Gallagher, the much-appreciated statistical advice of Dr. Andy Coldman, the considerable expertise and advice of Dr. John D. Potter, and the kind permission to use the data given by the original principal investigator, Dr. Alice Whittemore.
GLOSSARY

Effective carbohydrate. Total carbohydrate minus fiber, or the portion of carbohydrate consumption that has an effect on blood glucose.

Gelatinization. An industrial process to improve palatability and digestibility by increasing branching of the starch polysaccharide; also increases glycemic index.

Glycemic Index. A ranking of foods based on the postprandial (following a meal) blood glucose response compared with a reference food, typically either glucose or white bread. Calculated as the integrated area under the curve of blood glucose response plotted over time.

Hyperinsulinemia. The state of having excess circulating insulin.

Insulin resistance. Loss of normal response to insulin by tissue receptors so glucose uptake is decreased.

Neutral detergent fiber (NDF). A group of polysaccharides made up mainly of cellulose and hemicellulose. NDF substances are generally insoluble and fibrous.

Pectin. A complex group of polysaccharides that are unbranched chains of 1,4-linked D-galacturonic acid units. Pectic substances are generally water soluble and gel forming.

Postprandial. Following a meal.
Infirmities of the stomach usually proceed from surfeiting.

Let such as have weak stomachs, avoid all sweet things, as honey, sugar, and the like; milk, cheese and all fat meats; let him not eat till he is hungry, nor drink before he is dry; let him avoid anger, sadness, much travel, and all fryed meats; let him not vomit by any means, nor eat when he is hot.

Nicolas Culpeper, 17th century physician and herbalist
INTRODUCTION

Purpose

The impact of dietary factors on the burden of cancer has not yet been accurately measured, but current best estimates attribute an average of 35% of all cancers to dietary causes, with a range of 10-70% (1). That average percentage makes human nutrition as relevant to cancer prevention as tobacco, and translates to over 22,000 deaths in Canada last year (2) which potentially could have been prevented. Globally, 3-4 million cases of cancer per year might be prevented by dietary means (3). The role of environmental factors in cancer is further supported by a recent large twin study which indicates environmental factors account for approximately 65% of colorectal cases (4). Colorectal cancer (CRC) has an intriguing 20-fold variation in international incidence (3). This magnitude of variation argues for environmental causes, especially when supported by evidence from studies showing that migrants from areas of low risk to areas of high risk take on the risk profile of their adopted country (5-7), sometimes in the same generation. For this reason, many environmental factors have been investigated with respect to colorectal cancer.

The present study looks at the risk of colorectal cancer associated with high levels of dietary carbohydrate consumption. In the effort to reduce fat and red meat consumption, North Americans often compensate by increasing their intake of refined carbohydrates,
thinking them to be harmless. If there are risks associated with increased carbohydrate consumption, it is crucial these risks be defined and quantified, because there are broad implications for cancer prevention strategy.

The research hypothesis of this study is that “total carbohydrate minus fiber” (hereafter called “effective carbohydrate”, that is, the digestible portion that stimulates insulin release) is a significant risk factor for colorectal cancer via chronic insulin stimulation. Given the protective role of fiber (8), traditional analysis of total carbohydrates might be expected to bias the result toward the null due to the embedded fiber, a negative confounder. Some of the conflicting results obtained to date in studies of dietary factors and cancer may be due to one or more uncontrolled variables. Effective carbohydrate consumption may be one of those variables, with a significant impact on health, and heretofore not properly controlled for in analyses using total carbohydrates.

Research Objectives

The specific objectives of this research study were to:

- **Produce Relative Risk Estimates.** Assess the relative risk of developing colorectal cancer for those consuming large amounts of effective carbohydrates, compared with those consuming small amounts, after adjusting for other dietary and non-dietary potential confounding variables.

- **Provide Future Direction for Prevention Research.** Generate new testable hypotheses for future research on the carbohydrate effect, by examining the data by gender, cancer site, age, years in North America and menopausal status.
Rationale

The study rationale is based partly on a preliminary feasibility study which indicated a significantly increased risk of CRC with increasing effective carbohydrate consumption, and partly on recent published work in the peer-reviewed literature. Conceptually, high intake of carbohydrates can result in too much circulating insulin, a known growth factor for colon cells. When little is known about a subject, such as effective carbohydrates and their association with colorectal cancer, the best and most prudent design can be a secondary analysis using data from studies of large populations done for similar objectives. The present study utilizes case-control data that were collected to study other dietary factors and colorectal cancer, providing a cost-effective method of assessing the feasibility of a larger more focused study. This way, sufficient sample size and power are attained without incurring new costs until the effect is confirmed. The data from Whittemore et al. (6) are comprehensive in terms of nutrient data, demographics, family and medical history, and reproductive variables, so these data are relevant to address the research objectives of the current study. Despite issues of recall and selection bias, case-control data were selected to maximize variation in consumption, an area where cohort study data often suffer. In dietary studies, where effect sizes are likely to be modest, a case-control design is more likely to detect an effect if one exists, particularly a design involving a migrant population where environmental differences are expected to be large.
The Following Sections

Chapter 2 reviews the literature and summarizes the evidence for a carbohydrate role in colorectal carcinogenesis. First, the history, patterns and relationships of diet and cancer are reviewed, with special attention to the two geographic areas of interest in this study, China and North America (NA). Next, colorectal cancer is discussed, including dietary as well as non-dietary risk and protective factors. Specific areas of carbohydrate metabolism are then reviewed, followed by a section on insulin, hyperinsulinemia and insulin resistance as they pertain to the risk of colorectal cancer. Finally, the influence of reproductive factors on the colon is discussed. Chapter 3 contains the methodological details of the study design, case and control selection, data collection and analysis methods. Chapter 4 presents the study results, and Chapter 5 discusses some of the implications and limitations of the study, then suggests potential mechanisms and links the results with the existing body of published literature. Chapter 6 provides a summary, then looks forward to future designs required to begin answering the remaining questions.
Chapter 2

BACKGROUND

Overview

In studies of dietary factors and colorectal cancer, epidemiologic, animal and intervention studies have often produced conflicting results, due in part to the known limitations of dietary studies. One or more uncontrolled risk or protective factors, however, might also explain conflicting results—i.e., confounding.

Recently, a unifying hypothesis has been proposed which suggests that obesity, inactivity, alcohol and the consumption of a typical Western diet are all associated with the development of insulin resistance and hyperinsulinemia, and that hyperinsulinemia may be stimulating growth of colorectal tumours (9-13). As yet unproved, several lines of research support a role for insulin in colorectal carcinogenesis (14-18).

Although a big carbohydrate dose would not result in abnormally elevated insulin levels in a healthy young person, decades of eating foods with a high glycemic index can result in a chronic state of elevated insulin (10), due to over-stimulation and the development of resistance by the insulin receptors. This resistance causes the body to produce more insulin to maintain normal blood glucose regulation. Insulin is a powerful master hormone and a known risk factor for several major diseases, including heart disease, non insulin-dependent diabetes (NIDDM), hypertension and cancer (10, 19). The following sections
will discuss diet and cancer generally, as well as carbohydrates and insulin specifically as they relate to possible causal pathways for colorectal cancer.

Diet and Cancer

Paleolithic Nutrition

The first truly modern human beings, *Homo sapiens sapiens*, appeared in the late Pleistocene, approximately 40,000 years ago. The genetic complement we share with those first humans evolved over millions of years. In the last 10,000 years, the period since the introduction of agriculture, very little has changed genetically. Unfortunately, however, *everything* has changed socially and culturally over that same period, less than 500 generations, resulting in a mismatch between the environment we are physically adapted for and the environment in which we now live (20). Humans diverged from chimpanzees over 7 million years ago, yet the human genome has changed by less than 2% in that time (21). So in a blink of time like 10,000 years, one would not expect to see adaptation for “recent” changes in environmental factors such as diet, especially changes related to chronic degenerative diseases. These diseases, although they account for approximately 75% of Western mortality (21), generally affect people after their reproductive years and are therefore not affected by evolutionary pressure. One such change with broad implications for health was the introduction of agriculture, replacing most hunter-gatherer societies. There are many resulting consequences, and this discussion will focus only on those related to dietary changes, particularly factors affecting the quality and quantity of carbohydrates available for consumption.
Pre-agricultural man lived as a hunter-gatherer. Prior to the introduction of agriculture, the diet was comprised of a wide variety of wild game, honey, fruit, uncultivated vegetable foods and to a very small extent, unrefined cereals. Today, a mere eight cereal grains provide 56% of the food energy and 50% of the protein consumed on earth (22). Access to high-carbohydrate foods for the hunter-gatherer was limited and seasonal, and these foods required substantial physical effort to obtain. Modern human physiological responses to carbohydrates are still adapted for scarcity, with the intention of giving a selective advantage to those with a condition called insulin resistance where glucose is shunted to the brain and placenta instead of being taken up by the muscles and other tissues. The modern metabolism, however, must now function in a context of year-round availability of abundant carbohydrates, and very low levels of physical activity, where strategies for scarcity may be out of place or even harmful.

The rise of agriculture did not occur simultaneously throughout the world — ranging between 5000 years ago (England, Scandinavia) and 10,000 years ago (Near East). After the introduction of agriculture, the percentage of meat in the diet declined, and plant foods came to dominate the diet, particularly cereals (grass seeds). This change had profound effects on average body size and general state of health. Early H. sapiens sapiens was an average 6 inches taller than descendants who lived after the development of agriculture, and post-agicultural skeletons show signs of protein and calorie deficiency and sub-optimal nutrition (21). Today, our diet has adequate protein as evidenced by increased average height (we are now almost as tall as our Neanderthal ancestors), but our diet is still very different from the Paleolithic diet for which our bodies are best adapted. The
potential mismatch between our bodies and the world we now live in may lie at the heart of many serious health problems.

![Diagram of Hunter-Gatherer Energy Sources vs Modern American Energy Sources]

Figure 1. Hunter-gatherer compared with modern Western dietary composition (Eaton, 1997).

Five main differences between modern and pre-agricultural diets are important for their potential health implications:

1. As shown in Figure 1, the hunter-gatherers of the Paleolithic ate more vegetables and fruit than modern humans do. Their estimated intake of vitamin C, calcium, iron and fiber was much greater than that of modern humans (22).
2. There are many "new" foods, introduced in the last 10,000 years, that now make up a large part of the diet, including dairy products, cereals, refined grain products, and potatoes. Many of these foods are carbohydrate-rich, energy-dense and nutrient-poor, acting by substitution to displace more nutritious foods from the diet. Some new foods, including wheat, contain compounds that act as anti-nutrients by blocking mineral absorption such as iron and zinc (23). These foods have been recently introduced into the diet, from an evolutionary perspective, and it might be expected that human physiology has not yet evolved mechanisms to deal with them efficiently. It is even possible that some physiologic traits that are adaptive in a pre-agricultural context, become harmful in the context of a modern diet (Neel 1997), such as insulin resistance.

3. Some foods that appear in both pre and post agricultural diets, such as meat and fat, are actually quite different although the same labels are used, perhaps misleadingly. The meat of Paleolithic times was very much leaner than the meat of today with a lower saturated fat content (today's meat is at least 25% fat compared with approx 4% fat for wild game) (24).

4. It is believed that Paleolithic man did not have the means to make alcoholic drinks. By contrast, modern Western man takes 7-10% of daily energy intake in the form of alcohol (21).

5. The pre-agricultural diet was characterized by variety, whereas agriculture itself, as well as the business that grew up around it, fosters mono cropping and monotony of dietary intake (this makes monetary if not dietary sense). Currently, of the
approximately 195,000 flowering plants with edible parts, a mere 17 plant species provide about 90% of mankind's food supply (22).

So, modern man eats less vegetable carbohydrate but has introduced significant quantities of refined grains, dairy products and alcohol to his diet, carbohydrates of quite a different nature, in addition to more fat. But, weren't Paleolithic humans less healthy than we are? After all, they certainly had a much shorter average life span. Maybe they just didn't live long enough to suffer from chronic degenerative diseases? There are two arguments against this. First, although the hunter-gatherer average life span was indeed shorter, there are individuals in modern hunter-gatherer societies that do live longer, into their 60's, and these individuals do not show signs of degenerative disease such as obesity, high blood pressure or evidence of coronary atherosclerosis (21). Second, modern youth often show early signs of degenerative diseases, while youth in hunter-gatherer societies do not (21). Although it is correct that trauma and infection killed many Paleolithic individuals before they could develop degenerative disease, the above two arguments support the low risk of degenerative disease in early man compared with modern man. Differences in diet may play an important role in that difference in risk.

The Post-Industrial Diet

The introduction of agriculture increased the quantity of carbohydrates consumed, but it was the Industrial Revolution, in the 18th - 19th centuries, that changed the quality of carbohydrates in the Western diet. The introduction of large roller mills to process grain resulted in a much more finely ground product, with most indigestible material such as
bran removed. The starch became more easily digested and gelatinized, giving it a much higher glycemic index (GI), or blood glucose response following a meal. The glycemic and insulin responses were two- to three-fold greater than with coarse ground or whole grain cereals (25). Potatoes, introduced in Europe at about the same time, also have a very high GI. Diabetes began to rise in Britain at this time (26), although sugar consumption had yet to start climbing. Modern hunter-gatherers, such as the Australian aborigines never adopted agriculture, and they exhibit high rates of NIDDM when exposed to a Western diet. Their insulin response to potatoes is twice that of Caucasians (27), perhaps due to less exposure.

In the last fifty years, a sedentary lifestyle and an alarming epidemic of obesity have added to the genetic background of insulin resistance, making the degree of hyperinsulinemia after a modern high GI meal much greater than that experienced by humans during the hunter-gatherer or even pre-industrial past. Today’s Western lifestyle includes increased consumption of highly processed, nutrient-poor and energy-dense convenience foods, which are often carbohydrate or carbohydrate and fat in composition. These filler or “junk” foods not only have effects on metabolism, they also displace more nutritious foods from the diet, and so can have dual adverse effects on health.

The many powerful actions of insulin have broad-ranging and serious consequences for health if chronic levels persist. Both the qualitative and quantitative changes in dietary carbohydrates as well as length of exposure to those changes may be predictors of diseases such as metabolic syndrome, heart disease, non insulin-dependent diabetes mellitus and cancer.
The Current Patterns of Diet and Cancer in China and North America

Today, cancer and the other major chronic degenerative diseases account for the majority of mortality in Western countries. Interestingly, there are other countries where diseases like cancer exhibit significantly reduced incidence and mortality. China is such a country, with lower incidence rates for many, but not all cancer sites. Breast, prostate and colorectal cancers are all less common, while stomach and oesophageal cancers are more common than in North America (3, 28). The patterns of dietary intake in China and North America differ in many ways, and each continues to change, making any comparisons over time difficult to interpret.

In China, as well as in Japan and Italy, recent increases in colorectal cancer incidence rates (28) demonstrate that even countries that had enjoyed low rates of colorectal cancer in the past are not immune to environmental changes that affect colorectal cancer risk. Among immigrants from China to North America, as shown in the original study data analysed here (6), risk increased with increasing years lived in North America.

North America exhibits quite a different pattern of cancer incidence from China, though environmental and cultural changes are beginning to attenuate the differences. In North America, high rates of prostate cancer, breast cancer and colorectal cancer contrast with low rates of oesophageal and stomach cancers. Except for lung cancer, which is high in both areas, China and North America generally have opposite patterns of cancer incidence. Within North America, significant differences can be seen when comparing across ethnic groups, such as the low rates of prostate cancer in Asian groups, compared with high rates among African-Americans (3).
In terms of diet, cereals are the staple food in most of Asia. In China, cereals make up about 69% of the Dietary Energy Supply (DES) (3). DES represents commodities as produced and does not correct for wastage, loss or inedible portions, so may only be a rough estimate of dietary intake on a population level. Intake of meat is low in China (about 8% of the DES), and varies with income. Asian diets contain few animal products, and the cereals (rice, wheat, millet and some corn) are supplemented with some legumes, vegetables and fruits. As countries develop economically, however, the proportion of cereals tends to decline while the proportion of fats and refined foods increases. For the time period 1960-1990, rapid cultural and economic changes brought about marked changes in diet throughout Asia. Intake of meat in China, for example, rose 342% (3) over that period, and consumption of saturated fats therefore increased as well. The rapid introduction of Western “fast” foods, such as Coke® and MacDonald’s® French fries, has and will continue to contribute to the changing diet and lifestyle patterns in China.

In the North American diet, no single food group predominates in the way cereals do in Asian countries (23% of DES compared with China’s 69%). Time trends (3) for the same 1960-1990 period show declines in North American consumption of fats (down 15-20%), and dairy products (down 20-30%), increases in vegetables and fruit (up 10-20%) and large increases in alcohol (up 30%).
Suspected Dietary Risk Factors

At one time or another, most dietary components have been examined with respect to their effects on health. During the first “Golden Age” of nutrition, in the early 20th century, great advances were made in understanding dietary deficiency-related diseases. Scurvy, beri-beri, rickets and Vitamin A-related blindness are now all preventable diseases, at least in the developed world where resources exist for prevention. During the last century, the developed world underwent an “epidemiologic transition”, whereby the diseases of poverty such as under-nutrition became less prevalent, and the diseases of affluence and aging such as cancer and heart disease became the major causes of death. Now, in what may one day be called the second “Golden Age” of nutrition, we have begun to examine seriously the effects of excess rather than deficiency. Of course, since equilibrium is generally the rule in physiologic systems, excesses in one area often result in deficiencies in other areas, so certain deficiencies by displacement must also be anticipated.

Some of the most-studied relationships in diet and cancer include associations between: fat (especially saturated fat) and risk of colon or breast cancer; fiber and protection from colon cancer; and vegetables and fruits and protection from lung cancer. With the exception of the vegetable and fruit/lung cancer studies, results have been disappointingly inconsistent. Animal, ecologic, case-control and cohort study designs provide conflicting results, leaving the reader to wonder if some crucial piece of information is missing. While the pitfalls of dietary studies are well known (the reader is referred to the 1998 edition of Willett’s book for a thorough coverage) (29), the large number of studies argues that some of the results are valid: the question is which ones?
The fat hypothesis has its origins in many excellent ecologic studies that documented a remarkable correlation between animal fat consumption by country and national colon and breast cancer incidence (30, 31). These studies are further supported by other observations at the population level, such as Seventh-Day Adventists who have low rates of both colon cancer and fat consumption (32). These results caused great excitement and resulted in many studies in different populations over the past two or three decades. Recent cohort studies and a combined analysis of 13 case-control studies (3) did not find clear evidence of a relationship between fat intake and CRC, as was observed in the ecologic studies. The ecologic results had predicted a strong association, but when you actually follow people over time, that association seems to disappear. This is an example of an ecologic fallacy, or failure of an expected ecologic effect estimate to reflect the biologic effect at the individual level. Members of cohorts, unfortunately, often have fairly similar consumption patterns, and do not exhibit the range of variation needed to detect an effect of moderate size. Now, a Canadian intervention trial is underway (33) to determine the preventative effects of a low-fat diet for women at high risk of breast cancer. The results are awaited with great interest.

Fiber and protection from colon cancer, as an hypothesis, had its modern beginnings with the observation by Burkitt that Africans with very low rates of colon cancer ate very high levels of dietary fiber (34). With fiber, a similar pattern has evolved as for dietary fat. Animal and ecologic studies show promise of a strong association, in this case protective, but this is not consistently borne out by the case-control studies or by prospective cohort studies. Intake of vegetable fiber was generally found to be protective, while intake of
grain (cereal) fiber was unrelated or associated with increased risk. A recent publication (35) reported no association in the Nurses' Health Study cohort, again leaving the reader to wonder what we're missing here.

Diet is a scientifically risky area to study because of the dual challenges of complexity and inter-relatedness of the dietary components. Humans consume literally thousands of biologically active compounds, only a few of which are currently measured routinely. Interdependencies and cross-reactions are rarely accounted for, and issues such as effects of food combinations, variations in composition due to growing conditions, or effects of cooking and processing methods are generally uncontrolled. A new complication is the advent of genetically modified foods, which may have effects that are not yet documented or controlled. Add to that the self-reported nature of dietary data, often for long-past time periods, and it is very challenging to sort out the relevant relationships.

Colorectal Cancer

Description of Sites

Colorectal cancer is actually a group of functionally related gastrointestinal (GI) cancer sites, including the right colon, the left colon and the rectum. Although they share many risk factors, the three sites have differences in their specific functions and their epidemiology. The right or proximal colon is itself made up of sub-sites; the cecum, the ascending colon, the hepatic flexure, the transverse colon and the splenic flexure. The right colon accepts newly released liquid chyme from the small intestine and one of its main functions is to re-absorb most of the bile acids and some of the water. The right colon acts as a fermentation chamber for bacteria, and is the only part of the colon that
must propel the contents upward, against gravity. Cancer of the right colon is more common in women (3). The left or distal colon is made up of the descending colon and the sigmoid colon, and its main function is water reabsorption. The rectum is made up of the rectosigmoid junction and the rectum, and functions to hold the waste material before it is evacuated. Distal colon and rectal cancers are more common among men (36).

The three sites share a common function of final processing for partly digested materials. The common exposures that result from performing that function represent clues to shared etiologic factors. The functional differences, however, suggest some areas where epidemiologic differences in incidence and distribution might be predicted and tested.

**Epidemiology**

Worldwide, colorectal cancer is the fourth most common cancer both in terms of incidence and mortality (3). Colorectal cancer is the second most common cancer for Canadian women, and the third most common for Canadian men (2). An intriguing characteristic of colorectal cancer is the large international variation in incidence, 20-fold from the areas with the lowest (India, China) to the areas with the highest incidence (North America, Australia) (3). Many factors differ, of course, between people living in China and people living in North America. This magnitude of international variation, however, does however provide opportunities for epidemiologic study, since variation in the study population is key when attempting to detect modest effect sizes such as those commonly associated with dietary factors.
Men and women get colon cancer at about the same rate, but men have 20-50% higher rates of rectal cancer (3). Among the colon cancer cases, women are found to have higher proportions of proximal (right) colon tumours. Right colon tumours have also been increasing in developed countries, while colorectal cancer overall is now increasing in previously low-risk developing countries (3, 28, 37). Colon and rectal cancers are often studied together as colorectal cancer because the risk factors appear to be similar for all sites (right colon, left colon and rectum).

**CRC Risk Factors**

Migrant studies and twin studies both suggest that colorectal cancer is determined more by environmental factors than by genetic factors (4-7), with up to 90% of colorectal cancers potentially attributable to environmental factors including diet (1). Among dietary factors suspected in colorectal carcinogenesis are saturated fat intake, consumption of red meat, alcohol and total caloric intake, while dietary fiber, anti-oxidant vitamins, calcium and vegetables have been associated with decreased risk (38). Additional non-dietary risk factors include obesity (39), lack of physical activity (6, 41) and family history (36). Inactivity is the factor most consistently associated with colon cancer risk, which is relevant to this study because activity may reduce circulating levels of insulin (9, 10). Alcohol and tobacco use may increase the risk of colon cancer, while long-term aspirin (or other non-steroidal anti-inflammatory drugs or NSAIDs) use has been associated with as much as a 50% reduction in risk (40). Women using hormone replacement therapy (HRT) or oral contraceptives (OC) have seen some protection from colon and rectal cancer (40-
but the changing nature of both formulations has created problems both with measurement of exposure, and with interpreting the data in light of today's exposures.

Until recently, carbohydrate consumption as a colorectal cancer risk factor has largely been ignored. Some studies have found sugar (39, 43) or starches (44-46) to be risk factors and with the proposal of an insulin-related mechanism, more support is accumulating (9-12, 47, 48) but the relationship is far from clear.

The Role of Carbohydrates

Definition of Carbohydrate Subgroups

Carbohydrate, like fat, is a diverse nutrient group, made up of several subgroups with quite different properties. Just as saturated fat is different structurally and functionally from monounsaturated fat, so too are monosaccharides, or simple sugars, different from pectin, a soluble type of carbohydrate fiber, yet both are called carbohydrates. This section will first describe the types of carbohydrates, then focus on the digestible carbohydrate groups, starches and sugars, and their effects on blood glucose and insulin levels.

There are many ways to categorize a group as diverse as carbohydrates, such as complex or simple, vegetable or grain-derived, polysaccharides or monosaccharides, each with a specific application. For this study, carbohydrates will first be divided into digestible and non-digestible groups as shown in Figure 2.
Digestible is assumed in this context to mean digestible by human digestive enzymes. Digestible carbohydrates are then further divided into sugars and starches. The sugars are a group of mono- and disaccharides such as glucose, fructose, lactose and sucrose among others. The starches are a group of long-chain glucose polysaccharides that on hydrolysis yield glucose monosaccharides. The starches have alpha-1,4 bonds that can be broken by human digestive enzymes (49).

Non-digestible carbohydrates, or non-starch polysaccharides (NSP) can be roughly divided into soluble (pectins and gums) and insoluble (neutral detergent fiber or NDF) groups. Non-starch polysaccharides are also long glucose chains, but they have beta-1,4 bonds, which are not broken by human digestive enzymes and the compounds are often referred to as fiber (50). The term fiber generally refers to a broad range of compounds,
some of them non-carbohydrates such as lectins. In this paper, fiber will be restricted to mean only carbohydrate fiber.

Although they are not digested in the small intestine, increasingly the sub-components of fiber are thought to have significant metabolic activity (50). The major part of fiber comes from plant cell walls, and any number of metabolically active bio-chemicals could be part of a given fragment of plant cell wall. It makes sense, therefore, that the range of potential metabolic activity might be large, and dependent on the specific plant as well as characteristics such as freshness. There is some evidence that the two types of fiber, pectin and NDF have distinctive physiologic roles (50). By treating all carbohydrates as the same, analyses of total carbohydrates may be misleading, introducing misclassification.

**Glycemic Index**

The carbohydrate types of interest in this study are the starches and sugars. This digestible portion of total carbohydrate will be referred to as “effective” carbohydrate, because it is the part of total carbohydrate that has an effect on blood glucose. Within this group, carbohydrate foods can be further classified by glycemic index (GI). GI is the blood glucose response following a meal and is defined as the integrated area under the curve of glucose concentration plotted over a fixed time interval after eating (51). The GI scale uses either glucose or white bread as the standard reference food, setting it equal to a GI of 100. Values for equal amounts of other foods, usually 100 grams, are then given in reference to that standard value. Using white bread, the most commonly used reference food, the following examples of some foods and their GI values will help to illustrate the
discussion. Each GI value is the mean of at least three studies to allow for variability in the results (51).

- Baked potato – 121
- White Bread – 100
- Sugar (sucrose) – 92
- Boiled new potato – 81

Many people are surprised to find that sucrose (granulated white table sugar) has less effect on blood sugar than bread or baked potatoes. Sucrose, however, is not made entirely of glucose the way starches are. It is a disaccharide of one molecule glucose and one molecule lactose. Lactose has a glycemic index of only 65 (51), so the impact on blood sugar is much less than the all-glucose product of starch hydrolysis. For starches, the glycemic response is a function of morphology of the starch granule, and degree of branching. There are two types of starch hydrolyzed by human digestive enzymes – amylose and amylopectin. By being more branched, amylose is more accessible to the enzyme amylase, and is digested more quickly, resulting in a higher GI value. Industrial processing (e.g., gelatinization) increases the branching to improve palatability and digestibility, increasing glycemic and insulinemic responses.

Another important point illustrated by the list of foods and their GI’s is the difference between young, high sugar produce and older high starch forms of the same produce, such
as new potatoes and baking potatoes. As plant material ages, sugars are converted to starch, resulting in a higher glycemic index when consumed.

Some additional factors that can affect the digestibility and glycemic index of a food include length of cooking, particle size (smaller sizes have higher GI's), and ripeness. Rice is an example of a food that varies widely in its GI, due to genetic differences in amylose content of the starch granules (51). The published GI tables, therefore, are only an approximation of the true value, given the many additional factors that are typically not measured or reported in dietary studies.

*Carbohydrates and Blood Insulin Levels*

When carbohydrates are eaten, various digestive enzymes hydrolyze the glycosidic bonds and release monosaccharides for use by the body as fuel. For example, sucrose is broken down to glucose and lactose by the enzyme sucrase. Starches are hydrolyzed by salivary and pancreatic amylases to maltose disaccharides, then to glucose by maltase action. The amylases are remarkably fast and efficient enzymes, which partly explains why starch can produce such a high glycemic index. The release of glucose into the bloodstream then causes a rise in blood sugar. How quickly the rise occurs can be affected by other food components such as fiber which slow absorption and reduce the blood glucose response of foods eaten at the same time. When blood sugar rises past a certain set point, the liver signals the beta cells of the pancreas to secrete insulin. Insulin functions to reduce blood glucose, which is toxic at high levels, thereby re-establishing equilibrium within the range of acceptable values. When blood sugar falls, the liver signals the alpha cells of the pancreas to secrete glucagon, which has the opposite effect, raising blood
glucose. Insulin and glucagon are always both present to some degree and their balance is critical to normal metabolism.

Insulin levels can also be affected without changes in blood glucose levels. Amino acids from the consumption of protein cause insulin levels to rise but do not elicit a commensurate glycemic response (52). Secretion of glucocorticoids, as in periods of stress will result in increased blood sugar, and a consequent risk in insulin secretion (53). Even the smell of food can cause insulin secretion. Cephalic phase reflexes, which are autonomic and endocrine responses, can be triggered by sensory stimulation such as the sight or smell of food, resulting in release of compounds such as insulin and gastric acid even before the food is eaten (54). Insulin is an anabolic hormone that ensures excess food is stored for use in times of scarcity by building cells and storing fat. Many of the problems of excess insulin are related to this normal growth function, triggered at the wrong time, in the wrong context, or in the wrong amount.

**Insulin, Insulin Resistance and Hyperinsulinemia**

*The Pathway from Carbohydrates to Hyperinsulinemia*

The digestion of carbohydrate foods described above does not necessarily lead to hyperinsulinemia. Normal insulin receptors are exquisitely sensitive and blood glucose can be regulated with minimal amounts of insulin. Other dietary behaviours such as eating low glycemic index, high fiber foods can also reduce the glucose response, thereby reducing the need for insulin secretion (10) As well, non-dietary behaviours such as physical activity can improve insulin sensitivity and so reduce the amount of insulin needed to maintain blood
glucose regulation (10). In a young active person, then, small amounts of insulin can control large amounts of dietary glucose intake satisfactorily without adverse side effects.

Many North Americans, however, are physically inactive and eat excessive amounts of both fats and high GI carbohydrates, all of which are determinants of insulin resistance and hyperinsulinemia (10, 55-57). A diet high in refined carbohydrates causes rapid intestinal absorption of glucose and postprandial (following a meal) hyperinsulinemia. Chronic high glycemic load, in the context of pre-existing insulin resistance (IR) can produce particularly high insulin levels because the muscle tissue is not taking up the extra glucose and more insulin must be produced to compensate. Some people are more insulin resistant than others, and an IR genotype is recognized, with an estimated prevalence of 35% (58). There are also groups of people with particularly high rates of insulin resistance, such as the Pima Indians or the Australian aborigines, perhaps due in part to their late introduction to agriculture and refined carbohydrates (27). Insulin resistance has a normal function and judging by its prevalence, must have been an adaptive trait until fairly recently in our evolutionary past. In the last 5,000-10,000 years, however, carbohydrates have become available in large quantities year-round. During most of human evolution, survival and reproductive success were improved by taking a "thrifty" approach and diverting available glucose to the areas that require it; the brain, the mammary glands and the placenta. Insulin resistance in other tissues may function to ensure scarce high-energy fuel gets to the high priority systems that are obligate glucose users and critical to species survival. Insulin resistance is a common feature of pregnancy today, preferentially sending glucose to the baby via the placenta (58).
In today’s high glycemic index carbohydrate environment, however, the “thrifty” IR genotype becomes maladaptive because there is now an excess, not a scarcity of glucose and insulin is now required on a regular basis to compensate and keep glucose below dangerous levels, a short-term imperative that has long-term consequences.

**Effects of Excess Insulin**

Insulin is a powerful anabolic hormone with a seemingly endless list of complex metabolic activities. In the proper context and amount, insulin is of course a necessary part of normal human metabolism. In excess, however, insulin can cause a number of undesirable conditions to occur. In 1988, Gerald Reaven put a name to the cluster of conditions associated with insulin resistance — “Syndrome X” (19). It is now better known as Metabolic Syndrome. The syndrome includes high blood pressure, hyperinsulinemia, impaired glucose tolerance (IGT), increased plasma triglyceride concentration, and decreased high-density lipoprotein (HDL) cholesterol concentration. Reaven concluded that insulin resistance and hyperinsulinemia are involved in the etiology of hypertension, non-insulin-dependent diabetes mellitus and coronary artery disease (CAD).

These consequences are in part the result of insulin’s normal blood glucose-lowering function, but either taken out of context or in excessive amounts. To reduce blood glucose concentrations, for example, insulin must either remove the glucose from circulation or add water to dilute it – in fact, insulin does both. Glucose is stored as glycogen first, then as fat. This process requires the construction of cells, so insulin turns on cholesterol production, adversely affecting lipid profiles. At the same time, to further lower glucose concentration in the blood by dilution, insulin can increase water retention...
by acting at the kidney and colon, driving up blood pressure as a side effect (59). So with insulin, “less is more” and any excess is at risk of causing unintended and unfavourable consequences.

Insulin and Colorectal Cancer

Insulin’s role in colorectal cancer is a fairly recent area of investigation, but the idea is consistent with insulin’s normal role as a growth factor for human colonic mucosal cells, which have both insulin and insulin-like growth factor I (IGF-I) receptors (10). In contrast to its blood glucose regulation role, insulin is believed to act through IGF-I receptors in its mitogenic effect on colonic carcinoma cells in vitro (14, 17). The mitogenic signal transduction may be mediated by p21ras (16) an important proto-oncogene in colon carcinogenesis. More support for insulin’s role comes from human prospective studies of subjects with non-insulin-dependent diabetes mellitus (an insulin-resistant state) who have an increased risk of colorectal cancer (15, 18). Two recent prospective studies designed to study heart disease also found as association between mean plasma glucose levels after glucose challenge and mortality from colorectal cancer 12-20 years later, although only small numbers of cases were observed (60, 61). Thus insulin, a known growth factor, is a biologically plausible agent in colorectal carcinogenesis, particularly at chronically elevated levels.

Insulin and Other Hormones

Hormones do not act alone and insulin is known to interact with many other compounds. Glucagon, the opposing metabolic hormone which acts to increase blood glucose has many points of interaction with insulin. Female reproductive hormones may
also have significant interactions with insulin, which has been investigated with respect to breast cancer risk (62). Insulin, especially in high concentrations is also known to react with IGF-I receptors. The colon has receptors for IGF-I as well as for insulin itself. Colonic cell turnover is a normal function of growth, and these IGF-I receptors may only present a problem under conditions of insulin excess when increased concentration results in unintended cross-reactions.

**Reproductive Factors**

*Reproductive Risk Factors for CRC*

Cancers usually described as hormone-sensitive, or considered to have hormonal etiologic factors include breast, prostate, ovarian and endometrial sites. Colorectal cancer has only recently been linked with reproductive factors. In 1980, McMichael and Potter (63) reviewed hormonal determinants of colon cancer. They outlined and subsequently continued to develop an hypothesis to explain the protective effects of exogenous hormone use and parity. The hypothesis concerns the effects of dietary factors on the production, degradation and excretion of bacterially produced secondary bile acids (64, 65). Recent case-control (66) and cohort (67) studies confirm exogenous hormone replacement therapy (HRT) and parity (68) as protective. In addition, studies have shown HRT to be protective for adenomatous polyps (69) suggesting an early mechanism of action, perhaps in cancer initiation. Slattery and colleagues (70) also observed that HRT improves survival of post-menopausal colon cancer cases, suggesting a different, later role, perhaps in slowing progression or in damage repair.
High-dose oral contraceptives (OC's), such as those commonly available in the 1960's and 1970's have also been shown to confer protection (63), possibly contributing to a noteworthy period effect whereby women in their 40's during the mid-1970's experienced an unexplained decrease in colorectal cancer risk.

Hormonal Interactions with the Colon

Sex hormones influence bile acid synthesis and the cholesterol:bile acid ratio (63). There are reciprocal variations in plasma cholesterol and bile cholesterol concentrations during the human menstrual cycle (64), and furthermore pregnancy results in a reduced bile acid pool, with increased lithogenicity (concentration of cholesterol in bile), which is thought to be protective. Bile lithogenicity is also increased by high dose endogenous estrogen, such as the early oral contraceptives. The colon is thought to have both estrogen and progesterone receptors (71), and there is evidence that most colon cancers arise in cells which have had the estrogen receptor (ER) silenced by hyper-methylation (36). It is not known why the ER protein is important to colonic epithelial cells or what downstream estrogen-responsive genes the loss might affect.

Clinical studies have shown that gastrointestinal symptoms vary with the menstrual cycle (71), and women frequently experience a change in bowel habit just before and during menstruation (72, 73). Insulin resistance occurs in pregnancy, sparing carbohydrates for the rapidly growing fetus (74) and cyclic changes in glucose tolerance and metabolism might also be expected from ovulation through the luteal phase of the menstrual cycle, in preparation for possible pregnancy. Women in North America experience many more menstrual cycles than Asian women do because they start earlier,
finish later, and are pregnant or lactating for less time. If this difference relates to the colorectal cancer risk of high levels of refined carbohydrates in the diet, it would help to explain the difference in colorectal cancer incidence between the women in the two geographic areas.

The Reversal at Menopause

Menopause results in many changes, in some cases complete reversals. One example that is relevant to the current study is an observed reversal of the relationship between body mass index (BMI) and risk of breast and endometrial cancer after menopause (75). Before menopause, a higher BMI is often associated with protection from risk of cancer. Huang and colleagues (76) also found the same for relative weight, adult weight gain and breast cancer risk. One hypothesized mechanism for the pre-menopausal protection is the suppression of ovulation in obese women. By having fewer ovulatory cycles, it is suggested that the women have less exposure to both estrogen and progesterone. After menopause, however, a higher BMI is often associated with increased risk of breast and endometrial cancers, perhaps related to the production of estrogen by adipose tissue, as well as increased estrogen availability due to decreased sex hormone binding globulin (SHBG) (77, 78).

Potischman et al. (75) evaluated the relationship of body mass and hormonal profile in pre- and post-menopausal endometrial cases and controls. They found that as BMI increased, total estradiol increased in post-menopausal women, but decreased in pre-menopausal women. The opposite role that obesity plays in cancer risk depending on
menopausal status may be related to differences in plasma estrogen concentration. This is potentially relevant to other hormonally sensitive cancers such as colorectal.

Summary

During the last 10,000 years, dietary carbohydrates have changed both in their nature and in their availability. Recently, North Americans have been striving to reduce their fat intake, which often results in reduced protein too because many protein foods have significant amounts of fat. So, people are consuming more carbohydrates as a percentage of total intake, and also in absolute terms, thinking this substitution practice to be harmless. These carbohydrates are mainly highly processed and refined forms, resulting in a chronic high glycemic load. The resulting overworked insulin response, layered on top of any genetic insulin insensitivity, physical inactivity, and obesity results in chronic high insulin levels and increased risk of several serious diseases, including colorectal cancer. Women may be especially at risk possibly due to variations in insulin resistance during pregnancy and during the menstrual cycle, in preparation for shunting glucose to the placenta and mammary glands for reproduction.
Chapter 3

METHODS

Study Design

The data come from a case-control study, published in 1990. The data collection methods are described in detail in the original publication (6). Briefly, data were collected during the study period 1981-1986 from Chinese men and women in North America and China. The study centres included Vancouver, San Francisco, Los Angeles, Ningbo and Hangzhou. The current study does not make use of the Chinese-resident data (Ningbo and Hangzhou) due to missing data on fiber consumption. In North America, there were 473 cases and 1192 controls in the study. The selection criteria were:

1. diagnosis of adenocarcinoma of the colon or rectum (ICD-O classification site codes 153.1-153.9 for colon and 154.0-154.1 for rectum) between January 1, 1981 and December 31, 1986 at age 20 or more years, and

2. both parents of Chinese ancestry, defined for cases as Chinese surname, or subject’s own assessment of ethnicity, or best guess of ethnicity from hospital admissions records, or birthplace. For controls, ethnicity is based on the subject’s assessment.

1 The data were generously provided by Richard P. Gallagher, one of the original investigators and Head of Cancer Control Research at the BC Cancer Agency.
All cases meeting the criteria were selected from the BC Cancer Registry, the San Francisco-Oakland Surveillance, Epidemiology, and End Results Program, and the Los Angeles Cancer Surveillance Program.

**Cases**

In North America, interview data were completed on 59% of the eligible cases. Of the 805 eligible North American patients, 235 died between selection and interview, 91 refused to participate, 58 were too ill or had moved, and 421 were interviewed. To assess potential bias from loss to interview by death, next of kin were contacted for those that died before being interviewed (n=235), and 52 relatives (22%) knew enough about the subject’s lifestyle and agreed to provide a surrogate interview. Data from these interviews showed no differences in mean values for colorectal cancer risk factors such as family history, physical activity, dietary, demographic, medical, reproductive and migrant factors so information from patient and surrogate interviews was pooled, giving data on 473 of the 805 eligible cases. Surrogate interviews were not performed for the controls, as virtually all of the eligible control subjects were interviewed.

**Controls**

Chinese-American population controls were frequency matched to cases by sex, age (at 5-year intervals), and residential location, in approximately 3:1 ratio. In Vancouver, controls were selected from the Medical Services Plan data, matched on school district of residence, while in San Francisco and Los Angeles, controls were recruited by house-
to-house canvassing of the case’s neighbourhood. Of 2219 potentially eligible controls, 446 could not be contacted because they were deceased or had moved, 581 refused to participate, and 1192 (54%) were interviewed.

**Interviews**

The questionnaire was administered by a trained professional interviewer in the person’s home, in the language of their choice. To ensure a common protocol, interviewers at all centres were trained by the same core group of trainers, and taped interviews were monitored throughout the study. During the study period, investigators from all centres convened at least twice a year to maintain communication and a consistent approach. The questionnaire covered diet and physical activity, as well as demographic, medical, body weight and height, reproductive and migrant factors such as length of residence in North America. The subjects were asked about themselves at age 21, age 40, and a reference year (the year before diagnosis for cases, and the year before interview for the controls).

**Diet History**

Participants reported their average frequency of consumption of 84 foods in 6 food groups: meat, fish and eggs; dairy; starches and sweets; vegetables; fruits; beverages. They also reported portion sizes for items consumed more than once a week. Food models were used to simulate Chinese and Western style foods. For mixed foods, participants helped themselves to simulated foods in their customary meat-to-vegetable ratio and each
part was quantified by measuring its volume and assuming 1 ounce of volume equals 1 ounce of food weight.

*Data Coding*

A common protocol was used for questionnaire coding and data entry. All nutrient intake variables were calculated by combining food frequencies with portion sizes and nutrient values obtained from three sources (79-81). Intake variables for saturated, monounsaturated and polyunsaturated fat were adjusted to reflect reported use of fats in cooking. A new variable was created for the current analysis, effective carbohydrate consumption (eCarb). The independent variable of interest, eCarb is defined as “total carbohydrate” minus “total fiber” and is measured in grams per day. Variables used from the original study data include; age (as a continuous variable), gender (0=male, 1=female), saturated, mono-unsaturated and poly-unsaturated fat consumption (grams per day), years in North America (0=0-9, 1=10-19, 2=20+), inactivity (hours spent sitting per day, self-reported), retinol consumption (mgs per day), calcium consumption (mg per day), fiber consumption grams per day of pectin and neutral detergent fiber) and education (0=LT 12 years, 1=12+ years). For women, variables for parity (ever/never), oral contraceptive use (ever/never), hormone replacement therapy use (ever/never), menopausal status and age at menopause were also used.
Data Preparation

The data were examined for missing or unusual values. Data on fiber were not recorded for the Chinese centres. Since the diet data are the focus of this analysis, only North American centres were included in the final data for analysis. During the bivariate analysis, 1013 cases were found to have missing data for the variable “sedentary occupation”. This very high rate of missing data was due primarily to the inability to classify activity levels of some occupations, thus this variable was dropped from further analysis. There were 3 cases with missing values for the variable for years in North America, but it was decided to retain this important variable, since the small number missing would not outweigh the strong predictive power of this variable. The selection of the variable to represent socio-economic status was also affected by missing values. The variable for income levels was missing 67 cases for males, and 78 cases for females. This was considered too much data to lose, so the variable for education level, which has complete data, was selected as a reasonable surrogate for socio-economic status (82, 83).

Summary Descriptive Statistics

Summary statistics were done using the SPSS Version 9.0 (84). Frequencies function, including means, standard deviations and tertiles of independent variables in the model, such as age, years in North America, carbohydrate consumption, saturated fat consumption and inactivity. All analyses were performed separately by gender. Independent T tests were used to examine whether statistically significant differences
exist between the cases and the controls in any of the suspected risk/protective factors being considered for the multivariate model.

**Bivariate Analyses**

Pearson correlations were run to examine the relationship of the outcome, CRC, to each explanatory variable separately. Each variable was run separately in a logistic regression model to get an idea of its association with CRC incidence, before going on to consider the possible role of confounders.

**The Multivariate Logistic Regression Model**

A multivariate logistic regression model was built for the final analysis, with CRC incidence as the outcome variable, effective carbohydrate intake as the variable of interest, and the variables listed above as potential confounders to control for.

Model 1: \[ \text{Logit (p), or \log (odds)} = \beta_0 + \beta_1X_1 + \beta_2X_2 + \ldots + \beta_nX_n \]

Where \( X_1 \) = effective carbohydrates, and \( X_2 - X_n \) are the remaining variables in the model.

The hypothesis to be tested is that this model (M1) is significantly better than the null model (M0) at predicting the probability of CRC. In this case the null model is one with all the variables except effective carbohydrate.
Variable Selection

Initially, variables of interest and potential confounders/known risk factors were selected from the research data. After the bivariate analysis, some variables were dropped, based on a non-significant Pearson's correlation with the outcome, or a non-significant change in the goodness of fit when added to a null model. The remaining variables were included in an initial model, then through iterative modelling, the final group of variables was selected to maximise the likelihood ratio and goodness of fit measures. In addition, some known risk factors were considered for inclusion in the model. For example, education was selected to represent socio-economic status. Total kilocalories consumed, another known risk factor, was not selected due to the presence of all macronutrients in the model, and the resulting statistical collinearity problem. Study centre was also analysed for effect before being excluded, to ensure that there were similar proportions across centres, allowing for pooling of the data.

Adjustment for Total Energy Intake

The “Energy Partition” or “Energy Decomposition” method (29) was used to approximate a nutrient measure that is independent of total energy intake, itself associated with risk of CRC. In this approach, separate terms for each macronutrient are entered into the multivariate model. This was then tested by adding the term for total kilocalories consumed, which was not significant and did not materially change the other coefficients. The energy partition approach was chosen because the macronutrients were of interest,
and would have been included, while the variable for total intake was not found to have an independent effect in these data.

*Model Building*

The SPSS Forward LR stepwise model building procedure was selected, which uses the Likelihood Ratio statistic. The results showed that the variables Protein and Beta-carotene could be dropped without affecting the fit. Beta-carotene therefore was dropped, but protein was kept because it represents a large portion of energy intake and is required for the total energy adjustment. This result was consistent with the Backward Stepwise model done next, which in addition dropped Pectin, a variable that was added back because it is part of the carbohydrate group. The final model was then run with all of the variables, using a chi-square statistic to test the hypothesis that the coefficients of the terms are non-zero.

*Model Checking - Goodness of Fit*

Goodness of fit was assessed by comparing the predictions to the observed outcomes using 2-by-2 tables and also by looking at how likely the results are, given the parameter estimates. Throughout the iterative modelling process, the change in the likelihood statistic was used as a guide to whether the new model was better than the previous one.

The Wald statistic was used to test the hypothesis that the variable beta coefficients were zero, to determine which variables were not contributing to the model, and these
variables were removed. For example the variable for Westernization index was dropped, due to collinearity with number of years in North America (Pearson's correlation coefficient of .398, p<0.01).

Model Checking — Outliers and Influence

Residual values were saved during the regression run, to allow checking for outliers and influential points. These were plotted against predicted values, and the resulting plots examined. Specific cases were then selected for examination. Two cases were found to have large leverage values (relative influence of an observation on the model's fit), indicating outliers that could affect the results. These subjects were both found to have large values for consumption of all food types, although not so large as to be a clear data entry error. Additional regression models were done excluding the two cases, but the differences were small and did not change the conclusion, so the two cases were left in the final model.

Model Checking — Interactions

All independent variables were tested for interaction with the variable of interest, eCarb, but only one, menopausal status was found to have a coefficient significantly different from zero at the p<0.1 level. Relationships tested include the interaction of effective carbohydrate consumption with saturated fat, monounsaturated fat, polyunsaturated fat, retinol, calcium, inactivity (hours sitting per day), time since migration
(years in North America), Quetelet's Index, level of education, and for the women, parity, age at menopause, HRT use and OC use.

Model Checking – Residuals

Standardized residuals (residuals divided by their standard deviations) were plotted against predicted values and showed a normal pattern, in that the largest residuals are for cases that have low intake (i.e. should be at low risk), and for controls with very high intake values. There were no residuals greater than 2.3 standard deviations, and the residuals were approximately normally distributed.

The Final Model

The final model includes 15 variables. The independent variable of interest, eCarb, was categorized into tertiles, based on gender-specific control population cut-points of "effective" carbohydrate consumption. The remaining independent variables were analyzed as continuous variables where the data were available, to conserve degrees of freedom in the model. Bonferroni adjustment for multiple comparisons was not performed, because the research was focused on only a few of the associations in the model, and classical multiple-inference procedures produce unnecessarily imprecise intervals (85). This decision increases the probability that some intervals have missed their target.
Final Model Checking

The final model was checked in the following ways:

• The residuals plot was rerun to check that the pattern was unchanged.

• The model was run with the interaction term that was closest to significance during iterative testing, Quetelet's Index, which only approached significance for men, but it was still non-significant and was not included.

No changes were made to the final model based on model checking.

Data Analysis

Logistic regression analyses were performed using SPSS software (84). The final model described above was used to produce odds ratios and 95% confidence intervals for males and females by anatomic cancer site (right colon, left colon, or rectum), with and without reproductive variables for the females. All P values are two tailed.
RESULTS

Overview

The results show a statistically significant increased risk of colorectal cancer from increased consumption of effective carbohydrates. In agreement with the original study, other highly significant predictors of CRC include age, saturated fat intake, calcium intake, retinol intake, family history and years in North America, especially if more than 20 years. Gender is also a significant factor, with females having a greater effective carbohydrate-associated risk overall, as well as a much greater site-specific risk in the right colon than males, and having considerable protection with the onset of menopause.

Descriptive Statistics

Table 1 compares cases and controls separately by gender, on selected demographic variables and other characteristics considered potential confounders in the current analysis (all tables and figures are at the end of the Results section). These data show that on average, CRC cases eat more of everything than controls do, except fiber and calcium, and have lived longer in North America than the controls. With respect to carbohydrates, both males and females consume more effective carbohydrate than controls, but the difference was only statistically significant among the women. Cases also eat less fiber of both types

References to “statistically significant” mean a p value less than or equal to 0.05, unless otherwise noted.
than controls, but only consumption of pectin by the men approaches statistical significance.

Both male and female cases are older than controls. Male cases have a mean age of 67.6 years compared with 66.0 years for male controls while female cases have a mean age of 63.9 years compared with 61.7 years for female controls, but these differences were not statistically significant at the 0.05 level. No significant difference was observed in body mass index (Quetelet's Index) between cases and controls of either gender. Level of inactivity as measured by hours sitting per day, however, was greater in cases than in controls, but was only statistically significant in the males.

In general, cases are less educated than controls, though this is only statistically significant, marginally, in the women (16.4% of female cases had more than 12 years of education compared with 23.9% of female controls). Family history is a significant factor, with 25.7% of male cases having a positive family history in a first-degree relative, and only 17.3% of controls. In women, 23.8% of cases had a positive family history, and only 15.6% of controls. Another important difference between cases and controls of both genders is the number of years they have lived in North America. Among men, 52.1% of cases have lived in North America more than 20 years, while only 36.2% of controls have been in North America that long. The difference in women is similar (40.7% compared with 31.2%) and also significant. The above potential confounders have been accounted for in further analyses.
Figure 3 shows the distribution by age group of effective carbohydrate consumption, the variable of interest in this analysis, for males and females. This graph highlights the marked gender difference in consumption, and supports the use of gender-specific tertile cut-points to measure the eCarb variable and stratification of data by gender during analysis.

Differences in Risk by Gender

Table 2 shows the confounder-adjusted odds ratios for gender-specific tertiles of eCarb consumption. OR's and 95% confidence intervals were produced by multi-variate logistic regression. The regression model controlled for known confounders: age (at diagnosis for cases, at interview for controls), Quetelet's Index (QI), years of education, family history, years in North America, inactivity (hours spent sitting per day), and nutrient consumption variables for the reference year (one year before diagnosis for cases, one year before interview for controls) including protein, saturated fat, monounsaturated fat, polyunsaturated fat, calcium, retinol, fiber (pectin and neutral detergent fiber). Both genders showed a dose-response effect, with a significant increasing risk from the lowest to the highest tertile of eCarb consumption. The males have a doubling of risk over that interval (OR=2.02, 95% CI=1.28, 3.19), while the females' risk is even higher (OR=2.54 and 95% CI = 1.44, 4.46). The linear test for trend was significant (p<0.01), supporting a dose-response relationship. From this perspective, the genders appear similar in their response to effective carbohydrate consumption.
Differences in Risk by Anatomic Site

The real surprise comes in Table 3 where the gender-specific risks, which looked fairly similar in overview, are split into the component cancer sites: right, or proximal colon, made up of the cecal, ascending and transverse colon; left or distal colon, made up of the descending and sigmoid colon; and the rectum. The females are at greatest risk in the right colon from increased eCarb consumption (OR=7.35, 95%CI=2.60, 20.80, highest tertile compared to lowest tertile), while the risk for the males is concentrated in the rectum (OR=2.78, 95%CI=1.40, 5.54). The females show a similar but smaller and less statistically significant risk in the rectum (OR=2.30, 95%CI=1.02, 5.18), but the males show no significant risk in the right colon. The men, however, show some risk approaching statistical significance in the left colon (OR=1.75, 95%CI=0.93, 3.31), while the women show no significant effect at all at that site. Figure 4 graphically shows the relationship of effective carbohydrate intake with colorectal cancer risk overall, while Figure 5 (a, b, c) shows the relationship for the specific sites, rectum, left colon and right colon respectively.

Differences in Risk by Age Group

When the data are stratified by age (Table 4), using the median age of 67 as the breakpoint, the risk is similar for women, whether they are above or below the median age. In the males, however, only those over age 67 are at risk from consumption of effective carbohydrates. The younger women, then, seem to experience the same level of risk from consumption of effective carbohydrates as the older women, though the reasons may be quite different, while the younger men are not affected in the same way as the older men.
For both genders, it is not until the third tertile that the risk becomes statistically significant.

**Differences in Risk by Years in North America**

When the data are stratified by years lived in North America (Table 5), using 0-9, 10-19 and 20+ categories, another surprising gender difference is observed. For women, it is the new immigrants (within the first 19 years) who are at higher risk (for highest tertile compared with lowest tertile; OR for 0-9 years in NA = 2.84, 95%CI=1.02, 7.90; OR for 10-19 years in NA = 3.11, 95%CI=1.05, 9.17; OR for 20+ years = 1.91, 95%CI=1.28, 6.58). For the men it is different, in North America at least 10 years before showing increased risk (for highest tertile compared with lowest tertile, OR for 0-9 years in NA = 0.96, 95%CI=0.41, 2.25; OR for 10-19 years in NA = 3.02, 95%CI=1.15, 7.93; OR for 20+ years in NA = 2.37, 95%CI=1.13, 4.99). As shown in Figure 6, both genders reach the highest risk after 10-19 years in North America, but the women show significant risk in the first 9 years, while the men do not. Both genders appear to experience some attenuation of risk after 20+ years. The decrease in risk for the women after 20+ years in North America could be due to a higher proportion of post-menopausal women, because the onset of menopause, as discussed in the next section, is a significant protective factor.

**Reproductive Factors in Women**

The addition of reproductive factors to the regression model for women increases the overall all-sites eCarb odds ratios and their statistical significance. Table 6 shows the
adjusted odds ratios by site for females, including data on the additional variables for age at menopause, parity ever/never, HRT use ever/never, and OC use ever/never. Several key points emerge from these data.

First, the risk in the right colon is increased by adjustment for the reproductive variables, with an OR of 10.70 (95% CI=3.34, 34.30) for the third tertile of eCarb intake compared with the first tertile of eCarb intake. The unadjusted odds ratio for this same interval was 7.35 (95%CI=2.60, 20.80). Second, the left colon risk is increased somewhat as well, from an OR of 1.41 to an OR of 1.80 for the third tertile of consumption compared with the first tertile. Third, the rectal risk shows the reverse effect and is decreased by the addition of the reproductive variables, from an OR of 2.30 to an OR of 1.98. Additionally the rectal cancer risk, which had been statistically significant, is no longer significant when the reproductive variables are included. The all-sites risk does not differ much after the reproductive variables are added, going from an OR of 2.54 to an OR of 2.61, so the site differences appear to cancel one another out once they are combined.

Independent effects of the reproductive variables differ. Menopausal status was observed to be a powerful and statistically significant protective factor for all sites, at any age and the younger the age at menopause, the greater the protection. Using pre-menopausal women as the reference group, an odds ratio of 0.17 (95%CI=0.06, 0.50) was observed for women experiencing menopause under age 45, vs. an odds ratio of 0.32 (95%CI=0.16, 0.64) for women experiencing menopause after age 50. Artificial
menopause results in an OR approximately mid-way between the age groupings, which may reflect the heterogeneous nature of this group of women.

The remaining reproductive variables are mainly protective as well. While not statistically significant, the results support previous studies in that OC use was protective (OR=0.71, 95%CI=0.37,1.37), HRT use was also protective (OR=0.32, 95%CI=0.07,1.50), and parous vs nulliparous status was protective (OR=0.50, 95%CI=0.22,1.15). Missing values for the HRT variable (N=10) resulted in smaller cell sizes and wider confidence intervals.

Menopausal status may exhibit effect modification of the eCarb risk, in that it may afford a greater protective effect for women who consume the highest level of eCarbs. The interaction term for menopausal status was the only interaction to approach statistical significance (p=0.07) of all independent variables tested. The interaction of age at menopause with eCarb consumption also approached significance and suggested that for women experiencing menopause before age 45, onset of menopause confers greater protection to high eCarb consumers (in the 3rd tertile), than for the moderate consumers (in the 2nd tertile) or low consumers (in the 1st tertile). For women experiencing menopause at later ages, the protection is similar for all levels of eCarb consumption. This observation suggests a changing relationship of reproductive factors and carbohydrate metabolism with age. A significant interaction term is best interpreted by stratifying on the effect modifier variable. Table 7 shows OR's for pre- and post-menopausal groups for all cancer sites. Sample size does not permit further stratification by cancer site. The data show a greater risk from eCarb consumption for the pre-menopausal women (OR for
highest tertile compared with lowest = 3.81, 95%CI=0.77, 18.94) compared with postmenopausal women (OR= 2.26, 95%CI=1.19, 4.32).
<table>
<thead>
<tr>
<th>Consumption Variables</th>
<th>MALES Cases (n=284)</th>
<th>Controls (n=698)</th>
<th>p-Value</th>
<th>FEMALES Cases (n=189)</th>
<th>Controls (n=494)</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effective Carbohydrates (g/day)</td>
<td>242.1 (95.9)</td>
<td>237.0 (90.0)</td>
<td>ns</td>
<td>208.5 (83.8)</td>
<td>196.3 (69.8)</td>
<td>0.05</td>
</tr>
<tr>
<td>Fiber - NDF² (g/day)</td>
<td>9.5 (5.5)</td>
<td>9.7 (5.8)</td>
<td>ns</td>
<td>8.5 (5.0)</td>
<td>8.7 (4.8)</td>
<td>ns</td>
</tr>
<tr>
<td>Fiber - Pectin (g/day)</td>
<td>3.2 (1.6)</td>
<td>3.4 (1.7)</td>
<td>0.06</td>
<td>3.1 (1.7)</td>
<td>3.2 (1.6)</td>
<td>ns</td>
</tr>
<tr>
<td>Calcium (mg/day)</td>
<td>584.5 (312.6)</td>
<td>585.7 (305.8)</td>
<td>ns</td>
<td>487.0 (315.0)</td>
<td>514.0 (302.0)</td>
<td>ns</td>
</tr>
<tr>
<td>Retinol (mg/day)</td>
<td>1934 (1446)</td>
<td>1611 (1308)</td>
<td>0.001</td>
<td>1564 (1159)</td>
<td>1372 (1022)</td>
<td>0.03</td>
</tr>
<tr>
<td>Protein (g/day)</td>
<td>104.4 (50.2)</td>
<td>93.7 (42.0)</td>
<td>0.002</td>
<td>84.7 (42.7)</td>
<td>77.9 (31.8)</td>
<td>0.05</td>
</tr>
<tr>
<td>Monounsaturated Fat (g/day)</td>
<td>51.1 (23.5)</td>
<td>44.6 (17.6)</td>
<td>&lt;0.001</td>
<td>40.9 (19.0)</td>
<td>37.6 (15.4)</td>
<td>0.03</td>
</tr>
<tr>
<td>Polyunsaturated Fat (g/day)</td>
<td>18.3 (6.9)</td>
<td>17.2 (5.9)</td>
<td>0.03</td>
<td>16.1 (6.8)</td>
<td>15.4 (5.6)</td>
<td>ns</td>
</tr>
<tr>
<td>Saturated Fat (g/day)</td>
<td>26.6 (16.0)</td>
<td>21.0 (10.9)</td>
<td>&lt;0.001</td>
<td>19.3 (12.0)</td>
<td>16.4 (8.2)</td>
<td>0.002</td>
</tr>
<tr>
<td>KCAL/day</td>
<td>2947 (1064)</td>
<td>2913 (816)</td>
<td>ns</td>
<td>2732 (1548)</td>
<td>2629 (1106)</td>
<td>ns</td>
</tr>
<tr>
<td>Personal Variables:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (yrs)</td>
<td>67.6 (12.2)</td>
<td>66.0 (12.2)</td>
<td>0.07</td>
<td>63.9 (14.5)</td>
<td>61.7 (14.9)</td>
<td>0.09</td>
</tr>
<tr>
<td>Quetelet's Index</td>
<td>2.5 (1.3)</td>
<td>2.4 (0.9)</td>
<td>ns</td>
<td>2.6 (1.8)</td>
<td>2.6 (1.8)</td>
<td>ns</td>
</tr>
<tr>
<td>Hrs Spent Sitting per day</td>
<td>8.8 (2.7)</td>
<td>8.1 (2.7)</td>
<td>0.001</td>
<td>8.3 (2.7)</td>
<td>8.0 (2.6)</td>
<td>ns</td>
</tr>
<tr>
<td>Demographic Variables (%):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-12 years</td>
<td>67.6%</td>
<td>62.2%</td>
<td>83.6%</td>
<td>76.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;12 years</td>
<td>32.4%</td>
<td>37.8%</td>
<td>16.4%</td>
<td>23.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family History (1st degree):</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>25.7%</td>
<td>17.3%</td>
<td>23.8%</td>
<td>15.6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>74.3%</td>
<td>82.7%</td>
<td>76.2%</td>
<td>84.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years in North America:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-9 years</td>
<td>22.7%</td>
<td>35.2%</td>
<td>28.6%</td>
<td>39.1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-19 years</td>
<td>25.2%</td>
<td>28.7%</td>
<td>30.7%</td>
<td>29.8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20+ years</td>
<td>52.1%</td>
<td>36.2%</td>
<td>40.7%</td>
<td>31.2%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Abbreviations: SD = Standard deviation, ns = not statistically significant at the p < 0.1 level.
2. P values(2-sided) calculated using 2 sample independent t test for continuous variables, Chi-square test for categorical variables.
3. Total Carbohydrate (not shown separately) is equal to Effective Carbohydrate plus Pectin and Neutral Detergent Fiber(NDF).

Table 1. Selected characteristics of colorectal cancer cases and controls by gender.
Figure 3. Mean eCarb intake by gender and age group.
<table>
<thead>
<tr>
<th></th>
<th>MALES</th>
<th>FEMALES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Sites</td>
<td>All Sites</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>OR²</td>
</tr>
<tr>
<td>Cases³</td>
<td>284</td>
<td></td>
</tr>
<tr>
<td>1st Tertile</td>
<td>92</td>
<td>1.00</td>
</tr>
<tr>
<td>2nd Tertile</td>
<td>91</td>
<td>1.39</td>
</tr>
<tr>
<td>3rd Tertile</td>
<td>101</td>
<td>2.02</td>
</tr>
</tbody>
</table>

1. Abbreviations used: N = Number of subjects; OR = Odds Ratio, CI = Confidence Interval, ns = not statistically significant at the p<0.1 level.
2. First tertile is reference category. All p-values are two-sided. Adjusted for age, education, family history, Quetelet's Index, years in North America, hours sitting per day (inactivity) and consumption of fat, protein, calcium, retinol and fiber.
3. Number of controls = 698 for Males, or approximately 232 in each tertile.
4. Number of controls = 494 for Females, or approximately 165 in each tertile.

Table 2. By gender: Association between tertile of effective carbohydrate consumption and colorectal cancer risk.
### MALES

<table>
<thead>
<tr>
<th></th>
<th>Right Colon</th>
<th></th>
<th></th>
<th></th>
<th>Left Colon</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Rectum</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>OR</td>
<td>95% CI</td>
<td>p-Value</td>
<td>N</td>
<td>OR</td>
<td>95% CI</td>
<td>p-Value</td>
<td>N</td>
<td>OR</td>
<td>95% CI</td>
<td>p-Value</td>
<td></td>
</tr>
<tr>
<td>Cases³</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Tertile</td>
<td>53</td>
<td>1.00</td>
<td>---</td>
<td>---</td>
<td>114</td>
<td>1.00</td>
<td>---</td>
<td>---</td>
<td>105</td>
<td>1.00</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>2nd Tertile</td>
<td>21</td>
<td>1.75</td>
<td>(0.79, 3.84)</td>
<td>ns</td>
<td>36</td>
<td>1.16</td>
<td>(0.68, 1.99)</td>
<td>ns</td>
<td>33</td>
<td>1.55</td>
<td>(0.86, 2.81)</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>3rd Tertile</td>
<td>17</td>
<td>1.51</td>
<td>(0.57, 4.02)</td>
<td>ns</td>
<td>38</td>
<td>1.75</td>
<td>(0.93, 3.31)</td>
<td>0.08</td>
<td>41</td>
<td>2.78</td>
<td>(1.40, 5.54)</td>
<td>0.004</td>
<td></td>
</tr>
</tbody>
</table>

### FEMALES

<p>| | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>OR</td>
<td>95% CI</td>
<td>p-Value</td>
<td>N</td>
<td>OR</td>
<td>95% CI</td>
<td>p-Value</td>
<td>N</td>
<td>OR</td>
<td>95% CI</td>
<td>p-Value</td>
<td></td>
</tr>
<tr>
<td>Cases⁴</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Tertile</td>
<td>45</td>
<td>1.00</td>
<td>---</td>
<td>---</td>
<td>60</td>
<td>1.00</td>
<td>---</td>
<td>---</td>
<td>75</td>
<td>1.00</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>2nd Tertile</td>
<td>14</td>
<td>2.54</td>
<td>(0.98, 6.57)</td>
<td>0.05</td>
<td>17</td>
<td>0.91</td>
<td>(0.43, 1.92)</td>
<td>ns</td>
<td>21</td>
<td>1.17</td>
<td>(0.57, 2.41)</td>
<td>ns</td>
<td></td>
</tr>
<tr>
<td>3rd Tertile</td>
<td>21</td>
<td>7.35</td>
<td>(2.60, 20.80)</td>
<td>0.0002</td>
<td>21</td>
<td>1.41</td>
<td>(0.60, 3.35)</td>
<td>ns</td>
<td>33</td>
<td>2.30</td>
<td>(1.02, 5.18)</td>
<td>0.04</td>
<td></td>
</tr>
</tbody>
</table>

1. Abbreviations used: N=Number of subjects; OR=Odds Ratio, CI=Confidence Interval, ns = not statistically significant at the p<0.1 level.
2. First tertile is reference category. All p-values are two-sided. Adjusted for age, education, family history, Quetelet's Index, years in North America, hours spent sitting per day (inactivity) and consumption of fat, protein, calcium, retinol and fiber.
3. Twelve male cases missing subsite indicators. Number of controls = 698 for males, or approximately 232 per tertile.
4. Nine female cases missing subsite indicator. Number of controls =494 for females, or approximately 165 per tertile.

Table 3. By site and gender: Association between tertile of effective carbohydrate consumption and colorectal cancer risk.
Figure 4. CRC risk by tertile of eCarb intake and gender.
Figure 5. Risk by tertile of eCarb intake for a) rectal, b) left colon, c) right colon cancer.
### MALES

<table>
<thead>
<tr>
<th></th>
<th>&lt; 67 years of age</th>
<th></th>
<th>67+ years of age</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>OR</td>
<td>95% CI</td>
<td>p-Value</td>
</tr>
<tr>
<td><strong>Cases</strong></td>
<td>108</td>
<td>1.00</td>
<td>--</td>
<td>ns</td>
</tr>
<tr>
<td>1st Tertile</td>
<td>29</td>
<td>1.49 (0.77, 2.90)</td>
<td>ns</td>
<td>63</td>
</tr>
<tr>
<td>2nd Tertile</td>
<td>37</td>
<td>1.45 (0.66, 3.18)</td>
<td>ns</td>
<td>54</td>
</tr>
<tr>
<td>3rd Tertile</td>
<td>42</td>
<td>2.56 (1.40, 4.66)</td>
<td>0.002</td>
<td>59</td>
</tr>
</tbody>
</table>

### FEMALES

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>OR</th>
<th>95% CI</th>
<th>p-Value</th>
<th>N</th>
<th>OR</th>
<th>95% CI</th>
<th>p-Value</th>
<th>N</th>
<th>OR</th>
<th>95% CI</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cases</strong></td>
<td>91</td>
<td>1.00</td>
<td>--</td>
<td>ns</td>
<td>98</td>
<td>1.00</td>
<td>--</td>
<td>ns</td>
<td>189</td>
<td>1.00</td>
<td>--</td>
<td>ns</td>
</tr>
<tr>
<td>1st Tertile</td>
<td>25</td>
<td>1.00</td>
<td>--</td>
<td>ns</td>
<td>31</td>
<td>1.00</td>
<td>--</td>
<td>ns</td>
<td>56</td>
<td>1.00</td>
<td>--</td>
<td>ns</td>
</tr>
<tr>
<td>2nd Tertile</td>
<td>26</td>
<td>1.85 (0.90, 3.79)</td>
<td>0.09</td>
<td>29</td>
<td>0.99 (0.50, 1.96)</td>
<td>ns</td>
<td>55</td>
<td>1.23 (0.76, 2.00)</td>
<td>ns</td>
<td>189</td>
<td>1.00</td>
<td>--</td>
</tr>
<tr>
<td>3rd Tertile</td>
<td>40</td>
<td>2.55 (1.12, 5.80)</td>
<td>0.03</td>
<td>38</td>
<td>2.90 (1.28, 6.58)</td>
<td>0.01</td>
<td>78</td>
<td>2.54 (1.44, 4.46)</td>
<td>0.001</td>
<td>189</td>
<td>1.00</td>
<td>--</td>
</tr>
</tbody>
</table>

1. Abbreviations used: N=Number of subjects; OR=Odds Ratio; CI=Confidence Interval, ns = not statistically significant at the p<0.1 level.
2. First tertile is reference category. All p-values are two-sided. Adjusted for years in North America, education, family history, Quetelet’s Index, hours spent sitting per day (inactivity) and consumption of fat, protein, calcium, retinol and fiber.
3. Number of controls = 698 for males, or approximately 232 per tertile.
4. Number of controls = 494 for females, or approximately 165 per tertile.

Table 4. By age group: Association between tertile of effective carbohydrate consumption and colorectal cancer risk by gender.
<table>
<thead>
<tr>
<th></th>
<th>0-9 years in NA</th>
<th>10-19 years in NA</th>
<th>20+ years in NA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N(^1) \text{ OR} (^2) \text{ 95% CI} \text{  p-Value}</td>
<td>N \text{ OR} \text{ 95% CI} \text{  p-Value}</td>
<td>N \text{ OR} \text{ 95% CI} \text{  p-Value}</td>
</tr>
<tr>
<td><strong>MALES</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cases(^3)</td>
<td>64</td>
<td>71</td>
<td>147</td>
</tr>
<tr>
<td>1st Tertile</td>
<td>14 1.00 ---</td>
<td>18 1.00 ---</td>
<td>59 1.00 ---</td>
</tr>
<tr>
<td>2nd Tertile</td>
<td>22 1.02 (0.45, 2.31) ns</td>
<td>21 1.38 (0.59, 3.22) ns</td>
<td>48 1.59 (0.92, 2.74) 0.09</td>
</tr>
<tr>
<td>3rd Tertile</td>
<td>28 0.96 (0.41, 2.25) ns</td>
<td>32 3.02 (1.15, 7.93) 0.02</td>
<td>40 2.37 (1.13, 4.99) 0.02</td>
</tr>
<tr>
<td><strong>FEMALES</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cases(^4)</td>
<td>54</td>
<td>58</td>
<td>77</td>
</tr>
<tr>
<td>1st Tertile</td>
<td>10 1.00 ---</td>
<td>20 1.00 ---</td>
<td>26 1.00 ---</td>
</tr>
<tr>
<td>2nd Tertile</td>
<td>12 1.07 (0.40, 2.86) ns</td>
<td>13 1.00 (0.37, 2.65) ns</td>
<td>30 1.72 (0.92, 2.74) ns</td>
</tr>
<tr>
<td>3rd Tertile</td>
<td>32 2.84 (1.02, 7.90) 0.05</td>
<td>25 3.11 (1.05, 9.17) 0.04</td>
<td>21 1.91 (1.28, 6.58) ns</td>
</tr>
</tbody>
</table>

1. Abbreviations used: N=Number of subjects; OR=Odds Ratio, CI=Confidence Interval, ns = not statistically significant at the p<0.1 level.
2. First tertile is reference category. All p-values are two-sided. Adjusted for age, education, family history, Quetelet's Index, hours spent sitting per day (inactivity) and consumption of fat, protein, calcium, retinol and fiber.
3. Number of controls = 698 for males, or approximately 232 per tertile.
4. Number of controls = 494 for females, or approximately 165 per tertile.

Table 5. By years in North America: Association between tertile of effective carbohydrate consumption and colorectal cancer risk by gender.
<table>
<thead>
<tr>
<th>FEMALES</th>
<th>Right Colon</th>
<th>Left Colon</th>
<th>Rectum</th>
<th>All Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N^1 OR^2 95% CI p-Value</td>
<td>N OR 95% CI p-Value</td>
<td>N OR 95% CI p-Value</td>
<td>N OR 95% CI p-Value</td>
</tr>
<tr>
<td>Cases</td>
<td>45 1.00 --- ---</td>
<td>60 1.00 --- ---</td>
<td>75 1.00 --- ---</td>
<td>180 1.00 --- ---</td>
</tr>
<tr>
<td>1st Tertile</td>
<td>10 --- ---</td>
<td>22 1.00 --- ---</td>
<td>21 1.00 --- ---</td>
<td>53 1.00 --- ---</td>
</tr>
<tr>
<td>2nd Tertile</td>
<td>14 3.43 (1.19, 9.87) 0.02</td>
<td>17 0.94 (0.42, 2.08) ns</td>
<td>21 1.10 (0.53, 2.30) ns</td>
<td>52 1.29 (0.77, 2.14) ns</td>
</tr>
<tr>
<td>3rd Tertile</td>
<td>21 10.70 (3.34, 34.30) 0.0001</td>
<td>21 1.80 (0.72, 4.54) ns</td>
<td>33 1.98 (0.86, 4.57) ns</td>
<td>75 2.61 (1.44, 4.72) 0.002</td>
</tr>
</tbody>
</table>

Menopause^4
- Pre-menopause | 9 1.00 --- --- | 20 1.00 --- --- | 21 1.00 --- --- | 49 1.00 --- --- |
- Art. menopause | 5 0.27 (0.06, 1.24) 0.09 | 5 0.10 (0.02, 0.39) 0.001 | 7 0.40 (0.11, 1.40) ns | 17 0.26 (0.11, 0.61) 0.002 |
- Nat. < age 45 | 2 0.14 (0.02, 0.98) 0.05 | 3 0.09 (0.02, 0.43) 0.003 | 2 0.17 (0.03, 1.01) 0.05 | 7 0.17 (0.06, 0.50) 0.001 |
- Nat. age 45-49 | 9 0.17 (0.04, 0.70) 0.01 | 7 0.07 (0.02, 0.25) 0.0001 | 14 0.42 (0.14, 1.25) ns | 30 0.25 (0.12, 0.52) 0.0002 |
- Nat. age 50+ | 19 0.20 (0.06, 0.74) 0.02 | 24 0.10 (0.04, 0.40) 0.0004 | 31 0.48 (0.17, 1.37) ns | 74 0.32 (0.16, 0.64) 0.001 |

HRT-ever vs. never | 2% 0.67 (0.07, 6.39) ns | 2% 0.75 (0.09, 6.63) ns | 0% na na na | 3% 0.32 (0.07, 1.50) ns |
OC-ever vs. never | 9% 1.19 (0.32, 4.40) ns | 13% 0.87 (0.33, 2.35) ns | 8% 0.46 (0.17, 1.25) ns | 14% 0.71 (0.37, 1.37) ns |
Parity-ever vs. never | 96% 0.57 (0.10, 3.24) ns | 93% 0.42 (0.12, 1.51) ns | 92% 0.54 (0.17, 1.71) ns | 95% 0.50 (0.22, 1.15) ns |

1. Abbreviations used: N=Number of subjects; OR=Odds Ratio, CI=Confidence Interval, ns = not statistically significant.
2. First tertile is reference category. All p-values are two-sided. Adjusted for age, education, family history, Quetelet's Index, hours spent sitting per day (inactivity), years in North America and consumption of fat, protein, calcium, retinol and fiber. Also adjusted for parity (ever/never), OC and HRT use (ever/never).
3. Number of controls = 494, approximately 165 in each tertile.
4. Pre-menopausal group is the reference category for menopause OR's.

Table 6. Female colorectal cancer risk by tertile of effective carbohydrate consumption, including reproductive variables.
Figure 6. CRC risk for 3rd tertile of eCarb intake by years in North America.
<table>
<thead>
<tr>
<th>FEMALES</th>
<th>Pre-menopause</th>
<th>Post-menopause</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N&lt;sup&gt;1&lt;/sup&gt;</td>
<td>OR&lt;sup&gt;2&lt;/sup&gt;</td>
<td>95% CI</td>
</tr>
<tr>
<td>Cases&lt;sup&gt;3&lt;/sup&gt;</td>
<td>52</td>
<td>1.00</td>
<td>---</td>
</tr>
<tr>
<td>1st Tertile</td>
<td>10</td>
<td>2.14</td>
<td>(0.53, 8.69)</td>
</tr>
<tr>
<td>2nd Tertile</td>
<td>16</td>
<td>3.81</td>
<td>(0.77, 18.94)</td>
</tr>
<tr>
<td>3rd Tertile</td>
<td>26</td>
<td>1.23</td>
<td>(0.76, 2.00)</td>
</tr>
</tbody>
</table>

1. Abbreviations used: N=Number of subjects; OR=Odds Ratio, CI=Confidence Interval, ns = not statistically significant at the p<0.1 level.
2. First tertile is reference category. All p-values are two-sided. Adjusted for age, education, family history, Quetelet's Index, years in North America, hours spent sitting per day (inactivity) and consumption of fat, protein, calcium, retinol and fiber.
3. Number of controls =494 for females, or approximately 165 per tertile.

Table 7. By menopausal status: Association between tertile of effective carbohydrate consumption and colorectal cancer risk.
Chapter 5

DISCUSSION

Overview

These results are preliminary and need to be confirmed, however the data do indicate more than a twofold increase in risk of colorectal cancer for those with a high intake of effective carbohydrates (third tertile) vs. those with a low intake (first tertile), after controlling for potential confounders including age, education, family history, inactivity, Quetelet’s Index, and consumption of fat, protein, fiber, calcium and retinol. This observed result gives support to the study hypothesis that increased consumption of effective carbohydrates is associated with increased risk of colorectal cancer.

When the individual cancer sites are analysed by gender, a marked difference in risk by site is revealed. The women are at substantial risk of right colon cancer (OR=7.35, 95%CI=2.60, 20.80) if they have a high eCarb intake, while men show no significant effect from eCarb consumption in the right colon at all. This result was not hypothesized a priori. The odds ratio for women increases to 10.70 (95%CI=3.34, 34.30) when the data are adjusted for reproductive variables including menopausal status, age at menopause, parity, oral contraceptive and hormone replacement therapy use. This increase may be simply due to the wider
confidence intervals caused by adding more variables to the model, or it could be due in part to removing the masking effect of an uncontrolled protective factor, menopause, which was not evenly distributed in the study population (pre-menopausal: 23.4%, post-menopausal: 76.6%).

In the left colon, men show a modest increase in risk by the highest tertile of intake (OR=1.75, 95%CI=0.93, 3.31) which approaches statistical significance. Women, however, show only a small and non-significant risk at this site from consumption of eCarbs (OR=1.41, 95%CI=0.60, 3.35).

For rectal cancer cases, both men and women show increased risk with increased eCarb consumption. For the men, the effect is stronger and more statistically significant (OR=2.78, 95% CI=1.40, 5.54) than for the women (OR=2.30, 95%CI=1.02, 5.18), but both are in the same direction and are of the same order of magnitude. The rectal effect of carbohydrate consumption is consistent with effects observed for alcohol consumption, especially beer consumption and rectal cancer. Since alcohol is a carbohydrate "food", some of the proposed mechanisms for alcohol's role in carcinogenesis may be applicable to these results.

Interesting temporal differences were observed between the genders with the men showing increased CRC risk with age while the women did not (using two age groups, <67 and 67+ years). This effect may be partly explained by the
protective effect of menopause in the women. The women were also at higher risk from high eCarb intake as new immigrants (0-9 years in NA) than the men. This difference suggests a greater sensitivity to dietary changes in the women of this population, or perhaps a faster acceptance of new foods. Both genders show a peak, then a small decrease in risk by 20+ years in NA, which may represent an initial enthusiastic and unconditional adoption of new lifestyles, followed by an eventual tempering and balancing of old and new ways.

The significant protective effect of menopause on the risk of colorectal cancer might seem to contradict Wu-Williams and colleagues' findings (68) of no association for menopause and colorectal cancer risk in this population. The effect, however, may be explained by the interaction of menopause and carbohydrate consumption rather than the menopausal status variable on its own. The addition of the interaction term causes the protective effect of menopause to decrease and to lose statistical significance.

Before discussing possible mechanisms for the observed effects, the influence of such factors as chance, bias and confounding must first be addressed and estimated, at least in terms of expected direction of effect.

External Validity

The study results must be interpreted with caution, and in full light of such issues as the selection of one ethnic group (Chinese) for the study population.
The use of a single ethnic group means that although the results may be applicable to other members of that same ethnic group, the generalizability to other ethnic groups is questionable. In this case, Chinese were specifically targeted in the original study because of the relatively low incidence of colorectal cancer among people living in China compared with people living in North America. It is possible that people of Chinese ancestry may be either more or less sensitive to the effects of dietary carbohydrates than people of other ethnic origins. To test this, it will be necessary to complete a new study with data from a multi-ethnic population. In the current study, this restriction was helpful because the clear demonstration of the effect in this one special group provides the rationale for studies of other more diverse groups.

Aside from the issue of using a single ethnic group, there are further threats to generalizability that complicate the study of dietary risk factors. Cultural patterns of dietary intake are changing rapidly worldwide, including China and North America. New convenience foods, effects of television advertising, and changes in what is considered acceptable and desirable have all had effects on the dietary choices available now compared with those of the 1980's when the data collection was carried out. Geographic areas formerly of low colorectal cancer risk like Asia are experiencing rapidly increasing incidence rates and the differences explored by this study may be disappearing, forcing new studies to increase their sample size to attain sufficient power.
Internal Validity

Confounding. The presence of uncontrolled potential confounding variables is a problem that affects all designs except for a well designed randomized controlled trial. In a case-control design, potential confounding variables can be controlled by matching in subject selection, by restriction as discussed above with respect to ethnicity, or by multivariate analysis if data were collected on the variable. In this study, there are at least three known potential confounders for which data were not collected – alcohol use, tobacco use, and non-steroid anti-inflammatory drug use (NSAIDs). All three have been implicated as possible factors in colorectal carcinogenesis, but the data are not available to control for their effects in this study. The direction and magnitude of the potential effects cannot be predicted. The absence of alcohol consumption data may not be an issue for the current study, however, given lower alcohol consumption among Chinese, especially among women. In a new study, data on these factors would be helpful to improve validity and confirm the results. Data on diabetes were also not available, but Type II diabetes (NIDDM) is hypothesized to be on the causal pathway, and so is not treated as a confounder. Type I diabetes would be a confounder, but the absence of these data is not a significant concern due to the low incidence (1.8/100,000 per year) in Chinese (86). Some residual confounding must always be assumed, because even the measured factors are not perfect measures.
Bias. The issue of dietary recall bias is an important one in case-control designs, because the protocol requires participants to recall their dietary intake from one or more years in the past, to avoid moderating effects of the cancer itself on dietary habits. People often are unaware of the nutrient content of the food they eat, so their reports must necessarily be based on calculations from self-reported food frequency data. Further, changes in dietary habits are rarely reported with accuracy, because these changes are not normally sharply demarcated in a person's life, but rather occur slowly over long periods of time. Recall errors generally result in essentially random misclassification unless cases and controls differ systematically in their recall errors. Such errors can mask the normally modest effect expected from dietary studies, resulting in inconsistent published data. In view of the careful design of the original study, the lack of evidence that cases and controls differed in their recall ability, and the fact that the current study found effects of reasonable magnitude (from 2 to 10 fold), recall errors did not completely obscure the effect.

Bias in selection of controls is a limitation of case-control designs. In this study, the use of population or neighborhood controls helps to reduce the possibility of other diseases that may affect dietary choices independently, as is a common problem with hospital controls. Population controls, however, typically exhibit a low response rate (54% in this case) and those who do respond may be more health-conscious or otherwise systematically different from members of the
general population. Another potential area for misclassification is the possibility of undiagnosed colorectal cancer among the controls, because no colonoscopy or other diagnostic procedure could be done. This is commonly encountered, due to the prohibitive cost of screening the control population. The result should be a bias of the effect towards the null, so given the magnitude of the observed effect, the bias in this study was not enough to attenuate the result.

Problems of temporal bias also apply to dietary case-control studies. The effect being studied may be a result of dietary intake during childhood, for example, and the relevant period of exposure must be correctly assessed, or the dietary information collected may simply be from the wrong time period, resulting in a null finding. In the current study, a significant effect was observed, which is consistent with migrant studies showing host country rates attained in the same generation (5-7). This contrasts with some cancer sites, where only intermediate incidence rates are attained, suggesting more of a role for childhood exposure.

For the analysis of female data and reproductive factors, there is an additional issue of bias relating to error in age at menopause. Menopausal status and age at menopause were observed to be significant factors affecting risk of colorectal cancer, but there are at least two issues of potential error for these variables. Error in recalled age at menopause increases with time since menopause (87).
Also, if women with simple hysterectomy are included in the data, the true age at menopause may be unidentifiable. Since the current study separated natural menopause from artificial (although simple hysterectomies were not identified within the artificial menopause group), the issue of simple hysterectomy would only affect the artificial menopause data. Error due to time since menopause, however, would affect all the data and bias relative risk values downward.

**Precision**

Aside from problems of bias as noted above, issues of precision must be addressed. Precision in measurement corresponds to the reduction of random error, so random chance must be considered as a possible explanation for the results. In the present study, the number of subjects and controls is large, and a statistically significant effect was observed, so random chance likely can be ruled out as the primary determinant of the observed effect.

Measurement error for key study variables may also affect the precision of a study. All measurements have error and different types of error have different impacts on epidemiologic results. There are four basic types of error.

Within-person random variation in dietary intake is an example of random error that can be corrected for by averaging many repeated measures to approach the true value. In the current study, however, repeated measurements were not
available. This type of random error would tend to cancel out in terms of direction in a large study such as this one.

Systematic within-person measurement error can occur if, for example, a food item or an instruction is misinterpreted by a subject, but not necessarily by all subjects. Repeated measures will not correct for this, so this type of error would require a second, or validation study.

Random between-person error can result from using only one or a few measurements per subject in the context of random within-person variation. The over and under-estimation error is expected to counterbalance, so the mean would still be the true mean, but the standard deviation would be expected to be larger, making the estimate less stable.

Systematic between-person error is the result of systematic within-person error that affects subjects non-randomly, causing the group mean to be incorrect. Using the example above, if a food item or an instruction is misinterpreted by cases but not by controls, this would affect all individuals, but not to the same degree because they may use the food to different degrees, and the direction of the error would be difficult to predict.

Dietary studies therefore have a number of issues that limit the validity of the results, most importantly the lack of practical methods to accurately measure diet.
Diet is a complex set of exposures that are strongly inter-correlated. The effect of one nutrient, for example, may depend on the level of another nutrient or an interaction that was not measured.

In addition, since virtually all individuals are exposed to most dietary factors, each exposure cannot be designated as present or absent, but rather must be treated as a continuous variable, which most often has a very limited range of values. The limited variation makes effects difficult to separate from the background "noise" (29, 85). This problem can be partially addressed by ensuring the highest level of variation possible among study participants, and by employing the best current measurement methods available.

Finally, there is the issue of recent improvements in the quality and quantity of food composition data now available for use in calculation of nutrient intake from food frequency data. To make use of these improvements, it would be helpful to compare the results using new food composition data with the results using nutrient values derived from the 1980's food composition data to see if significant differences would be observed.

**Parsimony**

Statistical modeling must always strive to balance parsimony with completeness to achieve a model with both accurate risk estimates and reasonably tight confidence intervals. Each variable must earn its way into the model as an
independent and significant effect on the dependent variable, as well as assessed regarding its relationship to the independent variable of interest - i.e., a potential confounder. Potential confounders that do not have a significant effect in the study population, determined via iterative modeling, can be removed to give smaller confidence intervals. In this study, the initial model was informed by the findings of the original study, which indicated that fat, physical inactivity, age, gender and years in North America were important with respect to colorectal cancer risk. From a review of the literature, these were confirmed, and additionally, family history (36), obesity (39), and red meat were suggested as risk factors, where calcium and fiber (38) showed protective effects. Due to the lack of food item data for this secondary analysis, protein was substituted for red meat. An indicator for socioeconomic status (SES) was also added to adjust for the known effects of SES on the incidence of metabolic syndrome (88). Education is a valid indicator of SES and its effects on cancer risk (82, 89) and was chosen as a surrogate due to missing values in the income and occupation data (83).

Factors that were removed to give a more parsimonious final model include study centre, which was not found to be a significant factor in this study and total kilocalories, which did not have a significant effect once the major macronutrients (fat, protein, carbohydrates) were controlled.
An even more parsimonious model was created and tested. The model did not contain education or body mass index. It also combined the two types of fiber into one variable, and combined the three types of fat into one variable. The confidence intervals were smaller (data not shown) but the value of controlling the additional potential confounders was considered greater than the improvement in confidence intervals, so the more comprehensive model was used.

A Proposed Biologic Pathway

After estimating the effects of chance and bias, and controlling for known confounders, a positive effect remains to be explained. The following part of the discussion is divided into three parts – 1) a proposed biologic pathway model; 2) a suggested mechanism for the overall increase in risk of colorectal cancer with increased eCarb consumption; and 3) suggested mechanisms for the specific increase in risk of right colon cancer in the women. Different mechanisms suggest themselves for the right colon effect in women compared with the overall effect in both genders.
Figure 7. Causal pathway.
A model of the biologic pathway proposed to explain some of the increase in risk of colorectal cancer is presented in Figure 7. The key component of the model is hyperinsulinemia, or the state of having excess circulating insulin. This state is a product of multiple interacting factors that can be categorized into two types: predisposing factors and contextual triggers.

Predisposing factors are those that directly influence insulin resistance, causing a rise in available blood glucose and a subsequent increase in blood insulin only if the glucose goes higher than a certain set point. Pregnancy, for example, often results in increased insulin resistance to ensure obligate glucose users like the placenta and the mammary glands have sufficient resources (58). In the context of scarce glucose, this is a sensible and adaptive process. Similarly, societies only recently introduced to a Western diet and abundant energy-rich foods, such as the Pima Indians, retain a higher than average proportion of the population with an insulin resistant genetic predisposition, adapted for their accustomed energy-dilute diet. Now, they are relatively hyperglycemic and hyperinsulinemic from a young age (10).

Contextual factors are equally important, as they determine whether an individual’s predisposition will be adaptive or maladaptive. The current North American culture is characterized by high intake of refined carbohydrates and low intake of soluble fiber, both of which contribute to a state of rapid intestinal
absorption and high post-prandial levels of glucose. In this context of plenty, mechanisms for hoarding glucose backfire, however good they were for survival in a different, glucose-scarce context.

Possible Mechanisms – Overall Effect

Insulin, a master anabolic hormone, is essential for human life. One mechanism that might explain some of the observed CRC risk is the potential for reaction with IGF-I receptors when insulin is available in excess amounts. In its normal range of roles, insulin-insulin receptor binding reactions occur in the context of a specific range of substrate concentrations. When insulin is present in abnormally high concentrations, however, it can react with other receptors, because receptor binding is a function of both affinity and concentration. One of the more infamous such cross-reacting receptors is the insulin-like growth factor (IGF-I) receptor. As the name implies, it has some affinity for insulin, which is further enhanced by increased insulin concentration. Also in the name is the primary role of IGF-I, which is a growth factor. Growth stimulated by the incorrect signal is not likely to be beneficial. Insulin’s mitogenic effect on colonic carcinoma cells in vitro may be via the IGF-I receptor (Watkins, 1990). Any cross-reaction with receptors intended for other purposes, especially cell growth, is a potential control problem since normal controls may not be operating in the abnormal context. Hyperinsulinemia is the key component of Reaven’s Syndrome X or Metabolic Syndrome, which is associated with numerous serious
conditions including hypertension, abdominal obesity, and an adverse lipid profile with increased triglycerides and decreased HDL cholesterol. This wide range of effects is consistent with a general, systematic fault such as cross-reaction with a number of receptor types under conditions of excess concentration. A “horizontal” mechanism, it applies broadly to possibly explain the many different types of symptoms observed.

Possible Mechanisms - Right Colon Effect

With respect to the right colon effect observed in female subjects, there are three possible mechanisms to consider. First, the cross-reaction with IGF-I receptors, discussed above, may be operating in the right colon. High carbohydrate intake and resulting hyperinsulinemia would lead to overstimulation of insulin receptors and insulin-like growth factor I receptors (IGF-IR) in colonic mucosal cells, a normal function initiated in the wrong context or in the wrong amount. There is no reason, however, to expect this mechanism would operate differently for women or for the right colon specifically.

Second, the effects on CRC risk of reproductive hormones is supported by data on oral contraceptives and hormone replacement therapy use as well as parity and menopause. Until quite recently, cancers of the colon and rectum had been exclusively classified as gastrointestinal (GI) cancers and their possible connection with the reproductive hormones largely unexplored. Fraumeni’s nuns
changed that by illustrating the increased risk of both colon and breast cancer associated with nulliparity. In the last three decades, evidence for a hormonal role in colorectal carcinogenesis has accumulated, informing two possible mechanisms, the bile-acid mechanism and the ovarian androgen excess mechanism, which will be discussed with respect to the current findings.

The bile-acid mechanism proposed by McMichael and Potter suggests that gender differences in bile-acid metabolic profile (65) are relevant to the preponderance of right colon tumours in women compared with men. With a higher pH and a higher ratio of secondary (bacterially degraded) to primary bile acids than men, women might be expected to have both a different bacterial population and a greater exposure to secondary bile acids. Some of the secondary bile acids, such as deoxycholic acid are suspected carcinogens (65). The site and gender specific differences in risk observed in the current data are consistent with this mechanism. The right colon in particular functions to reabsorb bile acids and acts as a fermentation chamber for intestinal bacteria. The liquid nature of the chyme and the uphill propulsion action means the right colon gets more exposure to luminal contents, increasing exposure to potential carcinogens. Direct evidence for a gender difference in bacterial population size and composition is provided by Lampe et al (91) in a randomized controlled trial (N=34) showing that fecal weight is greater for men than for women on the same intake, suggesting lower bacterial numbers in women. Lampe and colleagues also
found that women digest more NDF fiber than men, suggesting perhaps a
different balance of bacterial species, or maybe reflecting the longer transit time
often observed in women (72). One possible way that diet, and specifically
carbohydrate intake might interact with colonic bacteria and secondary bile acid
production is via the intermediary of female reproductive hormones.

With respect to the interaction of the colon and reproductive hormones, the
colon has receptors for estrogen and although the messages passing from the
ovary to the colon are not yet clear, there are plausible reasons for such lines of
communication. For example, a down-regulation of bacterial numbers after
ovulation might function to reduce the chance of infection and increase the odds
of survival for the infant, an adaptive response in an environment where children
are born outside of modern health care facilities. An increase in transit time,
which is already longer in women, might achieve this by providing additional
fermentation time, and reducing the eventual numbers of bacteria present.
Additional fermentation time would also result in more degradation of primary
bile acids to secondary bile acids, with potentially negative consequences for the
mother.

Bacterial metabolism and bile-acid production may be directly affected by
carbohydrate intake. A recent study (92) in South Africa has suggested that
dietary starch that escapes digestion in the small intestine may be more important
than dietary fiber for fermentation in the colon, resulting in production of more short chain fatty acids. A high starch and sucrose, low fiber diet would be expected to result in quite a different food source for the colonic bacteria than a high fiber, low starch and sucrose diet would, producing a different bacterial output in which the right colon is then bathed.

A third possible mechanism, ovarian androgen excess, is based on insulin's role in sex hormone production and proposes that insulin resistance (IR) can result in an hormonal shift, especially if experienced during adolescence, and is related to a hyper-androgenic profile, anovulation and PCOS (polycystic ovary syndrome) (13). Insulin is a known factor in steroidogenesis and high levels of insulin potentiate the luteinizing hormone-stimulated production of ovarian androgens, particularly testosterone. Insulin has also been found to inhibit the production of sex-hormone binding globulin (SHBG) (93) which would result in higher plasma concentrations of free estrogen and testosterone. There is some evidence that insulin may amplify or even mediate entirely its effect on steroid hormone metabolism by down-regulating IGF-BP1 concentrations and therefore increasing availability of biologically active IGF-I in the target organs such as the ovary and breast (94). Through an increase in androgenicity, decreased insulin sensitivity that may initially be associated with mild abdominal adiposity can result in more central body fat distribution, and by interaction with growth hormone (GH) may stimulate mobilization of free fatty acids (FFA) and reduced uptake of
glucose, which further elevates plasma insulin concentrations (13). Hyperinsulinemia, then, may play a key role in the development of a hyperandrogenic endocrine profile that may increase the risk of hormonally sensitive cancers.

CRC vs. Sub-sites

Colorectal cancer is often used as the unit of analysis, based on the working assumption that right colon, left colon and rectal cancers have similar etiologic factors. The epidemiology of the three sites is already known to differ (3), and by the results of the current study, the dietary risk factors may differ significantly as well. The simplifying assumption that they can be discussed or studied as a single entity must therefore be used with due caution. As gastrointestinal cancers, there are certainly common features, but it is their differences that provide the most valuable etiologic information, and by analyzing them together, there is a danger of misclassification that can obscure those enlightening differences.

Evaluating the Evidence for Causality

Hill (1965) described several, now classic criteria in establishing if an effect is causal. The strength of effect, consistency across studies, experimental evidence, dose-response relationship, temporality, biologic plausibility, coherence with existing data and ability to predict analogous testable hypotheses are all factors to be weighed when judging the evidence for causality. In this study, the strength of
the effect observed was sufficient to be convincing, in light of the modest relative risks normally obtained in dietary studies. The results are not consistent with all previous studies on carbohydrates, but this may be explained by the novel treatment of carbohydrates in this study, that is to subtract the fiber portion before analysis, something that other studies have not done. Experimental studies in animals contribute to the biologic plausibility, but experimental studies in humans are mainly limited to transit time studies, and are not yet a significant contributor to the case for causation. There was a dose-response relationship observed, but it was only significant in the higher intake levels. A dose-response relationship may not be valid over the entire range of intake, so generalizing this finding would be difficult without further data on a wider range of intake. As a criterion of causality, biologic plausibility is nice to have, but is not required because the lack of such support may simply be due to insufficient knowledge at the time rather than lack of plausible mechanisms. In this case, however, insulin provides a biologically plausible agent for the observed effect, even though the exact mechanism or mechanisms are not yet clear. The relationship of diet and cancer is a complex one, and coherence with existing data is a challenge for most studies due to the large body of conflicting results. This study supports the more recent work on carbohydrates and insulin resistance, but there is not yet sufficient work in the area to provide a coherent background for comparison. Some of the previous studies focused on sucrose consumption rather than all carbohydrates,
but given the biologic mechanism proposed here, sucrose consumption is only part of the problem, so by studying it in isolation from starch consumption, the likelihood of a null result is increased.

If insulin is a causative agent of colorectal cancer, some testable analogies would include: first, that diabetics would have a higher risk of colorectal cancer and there is some recent evidence for this (15, 18); second, that colorectal cancer cases would have higher plasma insulin levels and a higher level of undiagnosed NIDDM than the general population. Unfortunately, this latter prediction is difficult to test because the disease, plus any surgery and treatments, changes too many factors relating to diet and metabolism to allow accurate measurements to be taken. A third class of predictions might be to expect similar risk relationships with carbohydrate intake and other hormonally sensitive cancers such as breast cancer. Del Guidice and colleagues found increased levels of circulating insulin in women with premenopausal breast cancer in a recent case-control study (62). Bruning and colleagues found serum levels of C-peptide to be significantly higher in early breast cancer cases compared with both population controls and 'other-cancer' controls, implicating hyperinsulinemia as a significant and independent risk factor for breast cancer (95). Kaaks and colleagues have presented an hypothesis of hyperinsulinemia as the physiologic link between nutritional lifestyle factors and breast cancer risk, mediated by development of a hyperandrogenic endocrine profile, suggesting an association with polycystic
ovary syndrome, an insulin-resistant state (13). Etiology of pancreatic cancer might also be expected to be associated with abnormal glucose metabolism, as shown by Howe and Burch's review where 4 of 5 studies showed a positive relationship with carbohydrate consumption though not all were statistically significant (96). This is further supported by a recent prospective cohort study showing a positive relationship between post-load plasma glucose concentration and pancreatic cancer mortality (97). Finally, in light of insulin's water retention action, renal cell carcinoma risk may also be positively related to eCarb intake in a subset of cases. This is suggested by an international case-control study of obesity and weight cycling and their relationship with renal cancer as mediated by metabolic syndrome components such as increased circulating levels of androgens and IGF-I, and decreased insulin sensitivity (98).

Overall, the evidence suggests a good case for insulin as an agent of colorectal carcinogenesis. More research using blood insulin levels, both fasting and post-load, as the measures will be helpful in confirming these findings.
Summary

There are four main conclusions from this study. First, increasing eCarb consumption is associated with increasing risk of colorectal cancer in both males and females. The risk ratio associated with high levels of consumption compared with low levels is 2.0 for men and 2.5 for women when all cancer sites are considered together.

The second conclusion is that all cancer sites should not be considered together because from these data, it appears that the dietary etiologic factors may differ by site, as shown by the female right colon risk magnitude and significance compared with the left colon risk.

The third conclusion is that men and women differ with respect to the etiology of right colon cancers. Carbohydrate consumption may pose special risks for women.

The fourth conclusion is that menopause is a significant protective factor, and appears to interact with carbohydrate consumption such that pre-menopausal
women are at higher risk, and early age at menopause confers most protection to women consuming the highest levels of carbohydrates.

Implications

In future studies of colorectal cancer, the right colon, left colon and rectum should be analyzed separately wherever possible. Data must also be stratified on gender to ensure important gender differences are not obscured. The risks of increased carbohydrate consumption, if confirmed by further research, may have broad implications for cancer prevention. Further research is outlined in the next section.

In summary, the current dietary context of high glycemic index foods and abundant, rapidly absorbed glucose has turned one of Nature’s survival strategies against us. In advising North Americans to reduce their percentage of fat instead of their absolute intake, they are encouraged to replace those calories with carbohydrates since many protein foods contain significant amounts of fat. If those carbohydrates were in the form of fiber-rich, energy-dilute vegetables and fruits, that strategy would be helpful. Unfortunately, the carbohydrates substituted are much more likely to be fiber-depleted, highly processed and very rapidly absorbed forms such as white bread. By swinging the dietary composition pendulum too far in the direction of carbohydrates, other risks become evident such as the colorectal cancer risk reported here. Ultimately, the “single enemy”
argument aimed exclusively at dietary fat must be retired and we have to work to achieve balance instead. Each dietary component has an important role in human physiology, but in excess or if out of balance with the other components, most things can become harmful.

**Future Research Directions**

The results of this secondary analysis support the *a priori* hypothesis of a risk of CRC with increased effective carbohydrate consumption, and they also provide direction for new work on the gender-specific carbohydrate-associated risk in the female right colon. Opportunities for further research into both hypotheses are numerous, and the following is a brief description of some possibilities.

The underlying problem aggravated by high effective carbohydrate consumption is hypothesized to be insulin resistance and hyperinsulinemia. The biggest limitation of the current study was the lack of a direct insulin measure, and the use of effective carbohydrate intake as a rough surrogate. In a new design, the collection of biological samples would be an integral component so the risk factors of interest can be measured directly, thereby providing much stronger evidence.

In addition, the generalizability of the results could be improved by using a multi-cultural study population. The focus on Chinese was appropriate for the
purpose of the original study which was looking at migrants from areas of low CRC risk (China) to areas of high CRC risk (North America), but a new study would be better served by a more diverse population.

The food and food group analysis in a future study could include the glycemic index variable which is hypothesized to be more relevant than simple carbohydrate content. The use of the latest nutrient composition data will also help to improve the precision of the study.

To further investigate the right colon risk in women, a new study must also focus on issues such as accurate measurement of age at menopause, and details of dosages and combination therapy for OC/HRT history. An additional data element that should be collected is HRT route of administration, which may have an effect on insulin action (99). Last, sufficient numbers of female right colon cases would be helpful to allow more sub-site and sub-group analysis.

In summary, a new population-based multi-cultural case-control study including biological samples for insulin levels and food group analysis by glycemic index will provide a logical next step in determining the role of dietary and associated hormonal influences on the risk of colorectal cancer.
BIBLIOGRAPHY


http://www.bccancer.bc.ca/research/ccr/people/ghislop/project1.htm; 2000.


