Pattern of Fruit and Vegetable Intake and its Association with Long-term Diseases in Different Ethnic Groups of Canada

MASTER OF SCIENCE (2008) MCMASTER UNIVERSITY

(Statistics) Hamilton, Ontario, Canada

TITLE: Pattern of Fruit and Vegetable Intake and Its Association with Long-term Diseases in Different Ethnic Groups of Canada

AUTHOR: Tanvir Quadir

Department of Statistics

Jahangirnagar University

SUPERVISOR: Dr. Noori Akhtar-Danesh

NUMBER OF PAGES: 86

Pattern of Fruit and Vegetables Intake and its Association with the Long-term Diseases in Different Ethnic Groups of Canada

A Thesis Submitted to the School of Graduate Studies
in Partial Fulfillment
of the Requirements for the Degree of Master of Science (Statistics)

Supervisor:
Noori Akhtar-Danesh
Faculty of Health Sciences
School of Nursing & Dept. of Clinical Epidemiology and Biostatistics
McMaster University

by

Tanvir Quadir
Graduate Student (Statistics)
Student No.: 0502201
Department of Mathematics and Statistics
McMaster University

Contents

List of Tables

Abstract

This ethnicity-based research based on Canadian Community Health Survey (CCHS) 2.2 demonstrates that majority people of all ethnic groups in Canada do not consume adequate times/servings of fruit and vegetables per day which might pose a potential risk factor for long-term diseases. The term 'ethnicity' is defined in this analysis with 'common cultural traits' instead of 'common gene pool'. Hence, Canadian diverse population has been decomposed into 14 explicit cultural/racial groups in CCHS 2.2. Significant differences in the consumption of 5 or more times/servings of fruit and vegetables per day exist among Canadian ethnic groups. Southeast Asian, Aboriginal people of North America, West Asian, Korean, Japanese and Chinese ethnic groups tend to exhibit the lowest intake of 5 or more times/servings of fruit and vegetables per day. As CCHS 2.2 is a complex survey, logistic regressions with bootstrap weights have been run to delineate the association between fruit & vegetables, race/ethnicity, and long-term diseases. Ethnic people with lower intake of total fruit and vegetables, fruit, and green salad have reported themselves to be more susceptible to long-term health conditions and chronic conditions. Low consumption of carrot and other vegetables are found to be responsible for bowel disorder and intestinal ulcers along with the long-term and chronic health conditions. Aboriginal people of North America has the highest propensity to contract most of the long-term diseases among all ethnic groups as opposed to the While ethnic group which may have an strong association with their lowest Fruit and vegetables intake.

Chapter One:

General Introduction

1.1 Preamble

The ethno-cultural profile of Canada reflects that it has become increasingly multi-ethnic and multi-cultural. Immigration to Canada over the past 100 years has brought a drastic metamorphosis of the nation's ethnic and cultural composition. Hence the Canadian population's ethnic diversity and diet pattern is potentially becoming a substantive part of health research in Canada. Ethnic groups, populations that share common cultural and linguistic characteristics, are known to vary widely in their burden of cardiovascular diseases, cancers, and other major causes of death (1). The study of ethnicity, disease, and diet can therefore provide valuable insight into fundamental questions of disease etiology and help to guide public health measures. Emphasis on the benefits of high intakes of fruits and vegetables is based on 1) the usually tacit assumption that such diets are very likely to be low in fatty foods, 2) the many observational studies reporting the association of high intakes with decreased incidence rates of cardiovascular disease and cancer (2-4, 5-7), and 3) the presumptive protection afforded by the various antioxidants in many fruits and vegetables (6). Increased intake of fruits and vegetables may thus provide a defence against oxidative stress, a potential target for preventing cancer and cardiovascular disease. This research analyses the Canadian Community Health Survey, Cycle 2.2 data set, known as the most recent national level nutritional survey and depicts the varying pattern of fruit and vegetable consumption in different ethnic groups and its consequent implications with the long-term disease components and other socio-demographic variables viz. sex, age. *Section 1.2* details the rationale for explicating the fruit and vegetable intake patterns in different ethnic groups. *Section 1.3* is designed to present the explicit objectives of this research analysis. *Layout plan* concludes this chapter which embodies the following chapters' succinct layout.

1.2 Background

This study stems mainly from the two-fold reasons that can be best described as:

(i) The multi-ethnic Canadian population lacks a solid understanding of its fruit and vegetable intake status in different ethnic groups and thereby its consequences on health components as a complete national level nutritional survey has long been absent since 1973. Ethnicity issue poses this problem on the top priorities as immigration and intermarriage have shaped Canada with more diverse groups of ethnic, racial and cultural origins. This study will predominantly focus on comparing and contrasting fruit and vegetables intake (FVI) of different ethnic groups of Canada with special attention to their meeting the threshold of the 5-a-day of FVI as prescribed by the Dietary Recommended Intake committee (8).

(ii) The burdens of long-term diseases in different ethnic origins that reflect health status are yet to be properly addressed. Ethnicity-based research is important as it documents the rates of known risk factors for a disease, identifies new risk factors, provides us with clues regarding similarities and differences in disease causation, and allows us to define high risk populations for specific diseases. Thus it helps us to understand variations in responses to preventive strategies, medical therapies and health care utilization patterns, and most importantly, it leads to specific prevention strategies that are appropriately tailored to the major ethnic groups.

Prior to 1990's, there was no recurring population health survey in Canada. National Population Health Survey (NPHS) was initiated in 1994-95 and continued every 2 years as a longitudinal survey persistently collecting information from the same sample of people on their health status and the factors that can have influence on health. In the late 1990's, as part of the health information roadmap initiative, Health Canada, Canadian Institute for Health Information, and Statistics Canada launched the Canadian Community Health Survey to address two basic questions: (i) How healthy is Canada's health care system? (ii) How healthy are Canadians? The Canadian Community Health Survey (CCHS) was intended to provide timely, consistent, cross-sectional estimates of health determinants, health status, and health system utilization across Canada. The first cycle of CCHS is labelled as "general" that includes a sample of approximately 130,000 Canadians, large enough to allow data to be presented at the level of health regions within each province and the second cycle that has a total sample of approximately 35,000 reflects the "focused components" including mental health and well being (CCHS 1.2), nutrition (CCHS 2.2), healthy aging (CCHS 4.2). This research is hinged on CCHS 2.2 that comprises of two components: (i) 24-hour dietary recall component that collects information on all foods and beverages during 24-hour period of reference (ii) the general health component part includes physical activity, sedentary activity, self-reported and measured height and weight, vitamin and mineral supplements, fruit and vegetable consumption, chronic conditions, smoking, alcohol, food security, socio-demographic characteristics, labour force participation, income and administration.

Fruit and vegetables intake might be one of the potential determinants of health status and prevalent diseases which provides more insight to carry out this research. There is strong and consistent evidence that diets high in fruits and vegetables can protect against the development of many cancers and cardiovascular and other chronic diseases (5,6). Canadian Community Health Survey results indicated that 67.2% of men and 56.5% of women 12 years of age and older consume fruit and vegetables less than five times a day (9). Currently, intake among adults and children is well below the National Cancer Institute-recommended minimum of 5-servings each day. These gaps between recommended and actual dietary intake are even more pronounced among ethnic groups owing in part to socioeconomic factors, limited access to and availability of quality fruits and vegetables, and socio-cultural determinants, placing them at higher risk of some diet-related diseases.

An understanding of the determinants of food choice in different ethnic groups can be used to design effective nutrition interventions to increase fruit and vegetable consumption. One of the emphatic reasons that triggered the CCHS 2.2 nutrition survey was to provide estimates of dietary intake in terms of nutrients, food and food groups, dietary supplements for a representative sample of Canadians at provincial and national levels. Hence more comprehensive research analysis may be carried out, although not attempted here, to present the techniques of intake distribution estimations, the use of the software *SIDE* (Software for Intake Distribution Estimation), and to interpret the estimates at the ethnic group levels.

Ethnicity: A Complex Construct

The ethnicity information was carefully examined at the very outset of this study to reach the criteria for determining ethnic groups. Information on the ethnic origins of the population has been collected in the census since 1901. The reporting of ethnic origin, and subsequent interpretation of the results, has become increasingly complex due to a number of factors.

The concept of ethnicity is fluid and is probably the most complex concept measured in the census. Increasing intermarriage among various groups has led to an increase in the reporting of multiple origins, which has added to the complexity of the ethnic data. A follow-up survey to the 2001 Census, called the Ethnic Diversity Survey, provided additional information to allow a better understanding of how Canadians of different ethnic background interpret and report their ethnicity. Ethnic Diversity Survey 2002 provides some comprehensive groupings of ethnicity origins based on *ethnic ancestry* and *ethnic identity*. Nevertheless, numerous ethnicity study papers agree that Ethnicity can be best tapped by *ethnic identity* instead of *ethnic ancestry* (10). *Ethnic ancestry* in CCHS 2.2 pertains to the question "To which ethnic and cultural group(s) did your ancestors belong?" The 21 ethnic groups come under this question are:

(i) Canadian (ii) French (iii) English (iv) German (v) Scottish (vi) Irish (vii) Italian (viii) Ukrainian (ix) Dutch (x) Chinese (xi) Jewish (xii) Polish (xiii) Portuguese (xiv) South Asian (xv) Norwegian (xvi) Welsh (xvii) Swedish (xviii) North American Indian (xix) Metis (xx) Inuit (xxi) Other

The crux of the problem raised by '*ethnic ancestry*' is that many of participants reported themselves as member of multiethnic groups. Parents from different ethnic groups or soaring reporting rate of 'Canadian' have exacerbated this problem. In 2001 Census, 11.7 million people, or 39% of the total population reported "Canadian" as their ethnic origin, either alone or in combination with other origins. This was up from 1996, when 8.8 million people or 31% of the population did so. On the other hand, in CCHS 2.2, although there is significantly lower rate of reporting "Canadian" ethnic group, it has become difficult to categorize all ethnic members into explicit group as they belong to several classes. In contrast, *ethnic identity* which is respondent's self-identification of the cultural and racial groups that he or she belongs is based on the question "People living in Canada come from many different cultural and racial backgrounds. Are you (i) White? (ii) Chinese? (iii) South Asian? (iv) Black? (v) Filipino? (vi) Latin American? (vii) Southeast Asian? (viii) Arab? (ix) West Asian? (x) Japanese? (xi) Korean? (xii) Aboriginal? (xiii) Other-Specify?"

As a result, *ethnic identity* question is deemed as the potential ethnic group determinant and the pertinent ethnic groups are considered in the analysis.

1.3 Objectives

In CCHS-2.2 *Fruit and Vegetable Consumption* is captured through variables namely *Fruit Juice, Fruit (Not including Fruit Juice), Green Salad, Potato, Carrots, Vegetables (Other)*. The prime objective is to describe the ethnic distribution; estimate the mean frequency of fruit and vegetable intakes in all ethnic groups and proportion of ethnic groups having 5 servings of fruits and vegetables per day. In our statistical analysis, variables are described with proportions and means across ethnic groups. The sampling weights endorsed by Statistics Canada are used to calculate estimates of proportions and differences in proportions of ethnic groups consuming less than or more than 5 serving of fruit and vegetables per day. To test for the differences between different ethnic origins in consumption of fruit and vegetables intakes, differences among all proportions are conducted to single out the ethnic people having potentially lower or higher intake of fruits and vegetables. The potential effects of FVI are exemplified by examining its relation with some common long-term diseases. Logistic regression models are developed to test whether diseases are associated with fruit & vegetable intake after adjusting for ethnicity and common confounders such as age and sex. Because CCHS 2.2 is a multistage stratified cluster design i.e. a complex survey, bootstrap macros provided by Statistics Canada are used to estimate odds ratio and 95% confidence interval of odds ratios to account for the design effect.

1.4 Layout Plan

Chapter Two includes a detailed literature review of ethnicity-based research encompassing fruit and vegetables consumption and the burden of diseases in different ethnic groups. *Chapter Three* provides a brief account of the CCHS 2.2 survey design and contents, describes the methodology used in estimating the dietary differences and its association with long-term diseases among all ethnic groups and a comprehensive description of the selected variables and their coding-recoding in logistic regression analysis. *Chapter Four* presents the empirical results, while *Chapter Five* reviews the main findings and limitations of this study and future research recommendations.

Chapter Two:

Literature Review

2.1 Introduction

The paucity of studies of fruit and vegetable consumption in relation to overall health in different ethnic groups spurs more concentrated research on the fruit and vegetable intake issues. Racial and ethnic disparities are more pronounced in recent research topics as diverse groups of Canadian people possess distinct features in their food choice and life style (11). Scores of studies have been carried out. Some of them shed light on perceived benefits of, and barriers to, fruit and vegetable consumption (12), some explore their relation to cardiovascular and chronic diseases (5-7), many of them are concerned about the determinants of fruit and vegetable intakes (13-14) and a substantial number of studies focuses on the correlates of fruit and vegetable consumption (15-16). But rarely have studies assessed the dietary behaviour patterns, including fruit, juice, and vegetable intake, among the different ethnic groups of Canada. The fruit and vegetable intake is pondered as one of the important factors that affect largely the interplay between diseases and ethnicity. The aim of this section is to provide a comprehensive review of notable works that accounted for fruit and vegetable intakes, diseases, and ethnicity issues. A literature review on the topic of fruit and vegetable intake seems to lose much in terms of coherence and logic if the discussion does not touch on its measurement and validity aspects. A brief account of the works on the validity of the fruit and vegetable questionnaire and data sets is presented at the end of this chapter.

2.2 Fruit & Vegetables and Diseases

Claudio analyzed the data of the first half of cycle 1.1 of the Canadian Community Health Survey which is conceived of as the first-time available population-based information on fruit and vegetable consumption in several decades across Canada (11). The study focused on the associations between the frequency of fruit and vegetable consumption and physical activity, smoking, obesity and alcohol-dependence. The study came up with some important findings about the fruit and vegetable consumption pattern of Canadian population indicating that women consume fruit and vegetable more frequently than do men. When other factors are taken into account, the frequency of eating fruit and vegetables is positively related in both sexes to being physically active, not smoking and not being overweight, and in women, to not being alcohol-dependent.

Anand in her Ph.D dissertation entitled "Ethnicity and the determinants of cardiovascular disease among South Asians, Chinese, and Europeans" exhibits that significant differences in the cardiovascular mortality rate exist between Canadians of South Asian, Chinese and European origin (17). In addition to the study of the relationship of the classical cardiovascular risk factors to disease outcomes, the contribution to disease outcomes of selected "emerging" risk factors (e.g., markers of thrombosis, socio-economic, dietary, and psychological stress factors) was also studied. The major findings reported in the thesis include that South Asians in Canada have the greatest prevalence, Europeans have an intermediate prevalence, and the Chinese have the lowest prevalence of cardiovascular disease. The dietary differences between the groups included the Europeans higher consumption of calories compared to the Chinese who had the lowest, and the sources of calories. The South Asians and European consumed more saturated and trans-fats than did the Chinese, and relatively more carbohydrates, whereas the Chinese consumed more protein and unsaturated fats. This study focused on nutrients intake from various dietary sources that includes fruit and vegetables but didn't establish any explicit association between fruit and vegetables and long-term diseases.

A prospective study was conducted within the Odyssey Cohort, which included men and women who volunteered for two cohort studies (CLUE I and CLUE II) established in 1974 and 1989 in Washington County, Maryland (6). The two CLUE studies drew their names from the campaign slogan, "Give Us a Clue to Cancer and Heart Disease". The Washington County, Maryland prospective study examined the association of fruit, vegetable, and antioxidant intake with all-cause, cancer, and cardiovascular disease death and revealed that higher intake of fruits, vegetables, and antioxidants may help protect against oxidative damage, thus lowering cancer and cardiovascular disease risk. CLUE participants who donated a blood sample in 1974 and 1989 and completed a food frequency questionnaire in 1989 were included in the analysis. Genkinger et al. found the participants in the highest fifth of fruit and vegetable intake had a lower risk of all-cause, cancer, and cardiovascular disease mortality (6). They demonstrated that higher intake of cruciferous vegetables was associated with lower risk of all-cause mortality and no statistically significant associations were observed between dietary vitamin C, vitamin E, and beta-carotene intake and mortality.

The intake of vegetables and fruits has been thought to protect against breast cancer. To examine the relation between total and specific vegetable and fruit intake and the incidence of breast cancer, Peeters et al. administered a prospective study of 285,526 women between the ages of 25 and 70 years, participating in the European Prospective Investigation into Cancer and Nutrition (EPIC) study, recruited from 8 of the 10 participating countries (18). Participants completed a dietary questionnaire in 1992-1998 and were followed up for incidence of cancer until 2002. Over 1,486,402 person-years (median duration of follow-up, 5.4 years), 3659 invasive incident breast cancer cases were reported. No significant associations between vegetable or fruit intake and breast cancer risk were observed. Olsen et al. manifested the same finding in their research that the overall breast cancer rate is not associated with the intake of fruits and vegetables but added that fruits and vegetables intake differentially affects estrogen receptor negative and positive breast cancer incidence rates (19).

An elevated LDL (Low Density Lipoproteins)-cholesterol concentration is associated with an increased risk of cardiovascular disease (20,21). In a recent study Djousse et al. used data from 4466 adult participants of the National Heart, Lung, and Blood Institute (NHLBI) Family Heart Study to evaluate whether higher intakes of fruit and vegetables is inversely related to LDL concentrations in men and women, independent of other risk factors (22). Participants in this study were members of families from previously established population-based cohort studies: The Framingham Heart Study in Framingham, MA; the Atherosclerosis Risk in Communities Study cohorts in North Carolina and Minnesota; and the Utah Health Family Tree Study in Salt Lake City. The authors used a food-frequency questionnaire to assess fruit and vegetable intakes and regression models to estimate adjusted mean LDL according to fruit and vegetable consumption. Fruit and vegetable consumption was inversely related to LDL: in the categories 0-1.9, 2.0-2.9, 3.0-3.9, and ≥ 4 servings/d, multivariate-adjusted mean (95% CI) LDL concentrations were 3.36 (3.28, 3.44), 3.35 (3.27, 3.43), 3.26 (3.17, 3.35), and 3.17 (3.09, 3.25) mmol/L, respectively, for men (p for trend <0.0001) and 3.35 (3.26, 3.44), 3.22 (3.14, 3.30), 3.21 (3.13, 3.29), and 3.11 (3.04, 3.18) respectively, for women (p for trend <0.0001). This association also observed across categories of age, education, smoking status, physical activity. Exclusion of subjects with prevalent diabetes mellitus or coronary artery disease did not alter these results significantly.

In a similar type of study on lipids in Health and Disease, Adebawo et al. underscored the fact that hyper-lipidemia is a major risk factor in etiology of cardiovascular disease (7). Their work reported the association of local fruit and vegetables with cardiovascular risk factors in African hyper-intensive subjects in an 8 week study. Though there was no significant difference in the Body Mass Index and HDL-cholesterol at the end of the eight weeks, there were significant reductions ($p < 0.05$) in serum triglycerides, total serum cholesterol, and LDL – cholesterol.

Diets rich in fruits and vegetables have been of interest because of their potential health benefits against chronic diseases such as cardiovascular disease (CVD) and cancer. Rissanen et al. assessed the association of the dietary intake of a food group that includes fruits, berries and vegetables with all-cause, CVD-related and non-CVD-related mortality (23). The subjects were Finnish men aged 42-60 examined in 1984-1989 in the prospective Kuopio Ischaemic Heart Disease Risk Factor (KIHD) Study. The risk of all-cause and non-CVD-related deaths was studied in 2641 men and the risk of CVD-related death in 1950 men who had no history of CVD at baseline. During a mean follow-up time of 12.8 years, cardiovascular as well as non-cardiovascular and all-cause mortality were lower among men with the highest consumption of fruits, berries, and vegetables. Consequently, the findings of this work indicated that diets that are rich in plant-derived foods can promote longevity.

Johnsen et al. also recommended increased intake of fruit for abating the risk of ischemic stroke based on prospective cohort study of 54,506 men and women who were included in the Danish Diet, Cancer, and Health study from 1993 to 1997 (24).

A systematic review on vegetables, fruits, and carotenoids and the risk of cancer by Ziegler presents a succinct overview of the major works dealing with the association of vegetables, fruits and carotenoids with the risk of cancer (25). In general, this study demonstrates that low intake of vegetables, fruits, and carotenoids is consistently associated with increased risk of lung cancer in both prospective and retrospective studies. In addition, low levels of β-carotene in serum or plasma are associated with the subsequent development of lung cancer. However, the importance of other carotenoids and other constituents of vegetables and fruit have not been adequately explored. Both prospective and retrospective studies suggested that vegetables and fruit intake may be associated with reducing the risk of cancers of the mouth, pharynx, larynx, esophagus, stomach, colon, rectum, bladder, and cervix. But because of fewer studies and less consistency among studies, the epidemiologic evidence is at present less persuasive than for lung cancer.

Similarly, Riboli and Norat summarized evidence from case-control and prospective studies on fruit and vegetable intake and cancer risk with a meta-analytic approach (26). The analysis included case-control and cohort studies that reported on total vegetable and fruit intake and risk of cancer of several sites published in English between January 1973 and June 2001; and referenced in the MEDLINE database (National Library of Medicine, Washington, DC). Case-control studies overall report a significant reduction in the risks of cancers of the esophagus, lung, stomach, and colorectum associated with both fruit and vegetables; breast cancer is associated with vegetables but not with fruit; and bladder cancer is associated with fruit but not with vegetables. Prospective studies provide weaker evidence than do case-control studies of the association of fruit and vegetable consumption with reduced cancer risk. The authors argued that the discrepancies in the results may be related to recall and selection biases in case-control studies. In contrast, the association may have been underestimated in prospective studies because of the combined effects of imprecise dietary measurements and limited variability of dietary intakes within each cohort.

In a prospective study on fruit and vegetable intake and the risk of cataract in women, Christen et al. suggests that high intake of fruit and vegetables may have a modest protective effect on cataract (27). Fruit and vegetable intake was assessed at baseline in 1993 among 39,876 female health professionals with the use of a validated, semiquantitative food-frequency questionnaire. During an average of 10 years of follow-up, 2067 cataracts and 1315 cataract extractions were confirmed. Compared with women in the lowest quintile of fruit and vegetable intake, women with higher intake had modest 10-15% reduced risk of cataract (*p* for trend <0.05).

2.3 Fruit & Vegetables and Ethnicity

Racial and ethnic disparities in fruit and vegetable consumption scenario have received little attention in Canadian health research. The following discussion of literature review is on ethnicity and dietary pattern of different ethnic groups.

Being urged by the low fruit and vegetable consumption in the Netherlands, Velde et.al. spearheaded a cross sectional study on the children of Dutch origin and non-Western ethnic minority children in the Netherlands to determine their differences in fruit and vegetable intake and determinants of intake (28). The empirical analyses exhibited that ethnic minority girls ate fruits more frequently than Dutch girls and no differences in frequency of vegetables intake were found among boys. Ethnic differences were found for almost all potential determinants. The Dutch children reported lower scores on these determinants than the ethnic minority children, except for perceived self-efficacy and barriers to eat fruit and vegetables. Knowledge of recommendations and facilitating behaviours of the parents mediated the association between ethnicity and fruit consumption among girls.

Enormous studies ponder dietary patterns as important precursors of disease and good health. Based on a multiethnic cohort study administered among 195,298 participants in Hawaii and Los Angeles in 1993-1996, Park et al. estimated intakes of food pyramid groups from a quantitative Food Frequency Questionnaire (FFQ) for subjects of 5 ethnic groups (African American, Hawaiians, Japanese Americans, Latinos, and whites) (29). Three distinct dietary patterns, "Fat and Meat", "Vegetable", and "Fruit and Milk", were identified by exploratory factor analysis with a Varimax rotation and validated by confirmatory factor analysis. Age, gender, and ethnicity had relatively strong associations with dietary patterns. Current smokers showed a positive association with the Fat and Meat pattern and inverse associations with the Vegetables. Physical activity was positively associated with the Vegetables and Fruit and Milk patterns but not with the Fat and Meat pattern. These findings propped up the hypothesis that dietary patterns are influenced by interrelated socio-cultural, demographic, and other lifestyle factors and may be useful in investigations of diet-disease relations.

To appraise the effects of gender and ethnicity in recognizing patterns in child and adolescent consumption of fruit and vegetables, Reynolds et. al. analyzed the data set from the four school-based sites included Alabama, Georgia, Louisiana, and Minnesota (30). These sites were selected in 1991 as part of the initiative National Five-a-Day for Better Health Program triggered by National Cancer Institute to encourage Americans to eat five or more servings of fruit and vegetables everyday. The research showed that girls ate more fruit, more vegetables and more fruit and vegetables combined than boys at the Georgia site. Ethnicity was significant in two sites: In Georgia, African-Americans ate more fruit and more fruit and vegetables combined than European-Americans.; in Minnesota, Asian-American/Pacific Islanders and African-Americans ate more fruit than European-Americans, and European-Americans and African-Americans ate more vegetables than Asian-Americans. No significant effects were found at the Alabama or Louisiana sites.

Ethnic differences in the risk of cardiovascular disease (CVD) and type II diabetes have consistently been identified in numerous studies. Lindquist et al. explored whether dietary factors explain ethnic differences in serum lipids and insulin profiles in children, independent of body composition and social class background (31). The study sample included 95 African American and white children. Macronutrient and food group intakes were derived from three 24-hour recalls. Intake of fruit and vegetables was found to be significantly higher and dairy intake lower in African American than in white children after adjustment for social class and total energy intake. Several direct relations were observed between diet and insulin action: carbohydrate and fruit intakes were positively associated with insulin sensitivity, and vegetable intake was negatively associated with acute insulin response. The African American children in the study showed a greater disease risk than did the white children, even after body composition, social class background, and dietary patterns were adjusted for.

Haire-Joshu et al. carried out an exploratory study with a randomized dietary intervention trial among 1227 African-American women to examine how estimates of one's fruit and vegetable intake in childhood are related to 3 dietary behaviours: intake of fruits and vegetables, exposure to and preference for fruits and vegetables, and preference for trying new foods (32). Results revealed that estimates of one's vegetable intake as a child were significantly related to exposure and preference for both fruits and vegetables, trying of new foods, and intake of both fruits and vegetables in adulthood whereas fruit intake in childhood was not proved significant in affecting the dietary pattern of adulthood.

Many studies hinged on the African-American racial group as these ethnic people were found to have the highest incidence and mortality rates for certain cancers (e.g., prostate) and the highest rate of hypertension in the world (2,3). African American men have higher mortality rates from heart disease and obesity than other racial groups and are 1.7 times as likely as white men to develop diabetes (4,33,34). Although overall fruit and vegetable consumption had not changed in the United States in recent years, consumption among black men fell dramatically and they take fewer fruits and vegetables than other racial or ethnic groups. In pursuit of best psychosocial correlates of fruit and vegetable consumption Moser et al. observed that fruit consumption, among African-American men appeared to be motivated by perceived benefits and standards set by important people in their lives; vegetable consumption was a function of extrinsic rewards and preferences for high-calorie, fatty foods (16). The results suggested that communications to increase fruit and vegetable consumption should be crafted to reflect differences in sources of motivation for eating fruits versus eating vegetables.

In a study of life-course events and experiences in three ethnic groups (black non-Hispanics, white non-Hispanics, and Hispanics) of the north-eastern United States and their association with fruit and vegetable consumption, Devine et al. hypothesized life events and experiences to have a long-lasting influence on food choice (36). Given the definition of food choice trajectories as “a person’s persistent thoughts, feelings, strategies, and actions as she/he approaches food choice”, it is believed to be formed as people are initiated into family or ethnic food traditions. Black, Hispanic, and white respondents differed significantly in life-course experiences, family roles, socio-demographic characteristics, and place of birth. Explanatory models for fruit and vegetable consumption differed among ethnic groups and between fruits and vegetables. Among black respondents, college education was positively associated with fruit consumption; education and family roles contributed most to differences in fruit and vegetable consumption. Among Hispanic respondents, life-course experiences such as liking fruits and vegetables in youth, making dietary changes for health, and food skills were found positively associated with fruit and vegetable consumption. Among white respondents, socio-demographic characteristics, such as being married with a young child or single with no child and having a garden as an adult, were to be positively associated with fruit and vegetable consumption. Variability in daily consumption of total fruits and vegetables was significantly higher ($p<.01$) among Hispanics (mean (SD) = 4.3 (4.0)) and blacks (4.4 (3.4)) than among whites (4.0 (2.3)).

Li et al explicated trends in fruit and vegetable consumption among adults in 16 US states as part of the Behavioural Risk Factor Surveillance System, 1990-1996 (37). The study reveals that the proportion of adults who consumed fruits and vegetables at least five times daily was 19%, 22%, and 23% in 1990, 1994, and 1996, respectively. While the proportion increased among those with active leisure-time physical activities and normal weight, it remained almost the same among inactive people and dropped among the obese.

In a study on US Hispanic women, infants and children (WIC) population in 1998 due to their declining average consumption of daily servings, Cotugna and Fleming attempted to elucidate their attitudes, knowledge and practice towards fruit and vegetable consumption (38). Only 29 participants (28%) were aware of the Five-a-Day for Better Health program. Sixty-four women (62%) stated that they had heard about the recommended amount of fruits and vegetables that should be eaten daily for good health, but when they were asked what the number was, their mean response was 3.6 servings. Seventy-nine participants (77%) believed eating fruits and vegetables might reduce cancer risk. Twelve participants did not believe this to be true, and the remaining 12 said they did not know.

Maskarinec et al. conducted a cross-sectional study to investigate the relationship between dietary patterns and body mass index among 514 women on the island of Oahu, Hawaii with different ethnic backgrounds who completed a validated food-frequency questionnaire (39). The authors described dietary patterns using factor analysis in an ethnically mixed population of women with Caucasian, Asian and Native Hawaiian ancestry. After adjustment for daily energy intake, the "meat" pattern was positively associated with body mass index, whereas the other three patterns showed negative relationships to body mass index for vegetables, beans, and cold foods. The associations were similar in direction and magnitude for all ethnic groups.

Cullen et al. performed a study entitled "Ethnic differences in social correlates of diet" where Grade 4-6 students completed questionnaires in the classroom and their parents completed telephone interviews (40). Questionnaires on parent- and child-reported family and peer influences on children's fruit, juice and vegetable consumption were analyzed for ethnic group differences in responses. Analyses of variances across ethnic categories and χ^2 analysis of differences in ethnic group composition between clusters of scales were conducted. Few ethnic group differences were detected, suggesting substantial commonality among respondents.

2.4 Validity of Measurements: Fruit and Vegetables Intake

To address the question whether 'five-a-day' is an effective way of increasing fruit and vegetable intakes, Bingham et al. analysed the data derived from the European Prospective Investigation into Cancer and Nutrition (EPIC) study (41). 269 men and women were sampled from EPIC-Norfolk to participate in the study assessing fruit and vegetables intake. The results indicate that the average portion of all fruits and vegetables measured was 87g, close to the standard portion size of 80g used as the basis of 'five-a-day' recommendations. Women ate more fruit than did men but fewer vegetables, so the total amount of fruit and vegetables eaten by men and women was the same. High consumers of fruits and vegetables ($\geq$400g/day) ate them 5 times a day whilst low consumers (< 400 g/day) ate them less often (3 servings per day, $p < 0.01$). Portion size differed little between high and low consumers. The authors concluded that frequency of intake is more important than portion size when distinguishing between high and low consumption of fruits and vegetables. Therefore, to increase intakes, it was recommended that low consumers should eat fruit and vegetables more often which endorses the 'five-a-day' healthy eating message.

Food frequency questionnaires (FFQ) are typically used to collect dietary data in various study settings. Controversy exists regarding the relative contribution of portion information to FFQ sensitivity, validity, and reliability (42-45). Campbell et al. analyzed the study conducted as part of the Black Churches United for Better Health Project, a five-a-day for Better Health community study targeting rural Black adult members of 50 churches in 10 eastern North Carolina counties (46). The study showed that portion models mailed before the interview can be used with a brief, telephone-administered FFQ to survey a rural black population. Adjusting the FFQ results with portion information increased estimated total fruit and vegetable consumption by approximately two-thirds mean daily serving, a substantial amount compared with the targeted study outcome of a one-half daily serving increase. The results suggested that rural blacks in North Carolina might be consuming more fruits and vegetables than has been previously estimated in national surveys.

Pomerleau et al. raised the challenge of measuring global fruit and vegetable intakes which affects the estimation of burden of disease (47). The WHO recently conducted a comparative risk assessment (CRA) within its Global Burden of Disease 2000 Study to estimate the global health effect of low fruit and vegetable intake. The paper summarizes the methods used to obtain exposure data for the CRA and provides estimates of worldwide fruit and vegetable intakes. Intakes were derived from 26 national population-based surveys, complemented with food supply statistics. Assessing exposure levels for the CRA had major methodological limitations, particularly due to the lack of nationally representative intake data. The results showed mean intakes generally lower than current recommendations, with large variations among subregions. The research suggests that if the burden of disease attributable to dietary factors is to be assessed more accurately, more countries will have to assess the dietary intake of their populations using comparable methods.

Chapter Three:

Methodology and Data

3.1 Introduction

The purpose of this chapter is to highlight the contents of Canadian Community Health Survey (CCHS) and to outline the bootstrap methodologies used in estimation of FVI and its association with long-term diseases, to provide details on modelling in logistic regression.

3.2 CCHS 2.2 Contents

Why is Nutrition focused in CCHS 2.2?

CCHS 2.2 was primarily aimed to estimate the distribution of usual dietary intake in terms of foods, food groups, dietary supplements, nutrients and eating patterns among a representative sample of Canadians at national and provincial levels. Nutrition was elicited as the focus of CCHS 2.2 for two-fold reasons: (i) Nutrition Canada Survey was the only other national level nutrition survey conducted by Canadian government a long time ago (completed between 1970 and 1972); (ii) Series of provincial surveys conducted over a 10-year period failed to provide a clear understanding of the nutrient intakes of the Canadians as food habits and food supply change over time or to make meaningful comparisons among provinces as nutrients intake information was not available for all ages of participants of all provinces. Although data from the United States are sometimes used as a surrogate for Canada, differences in the food supply, ethno-cultural characteristics, food habits, and fortification practices between the two countries pose the question of validity of such proxy information. Thus availability of current Canadian nutrition data was a priority.

The Target Population and Response Rate of Survey

The target population consists of all individuals living in private dwellings in the 10 Canadian provinces. To ensure a minimum number of individuals in each of 15 age-sex groups (or Dietary Reference Intake (DRI) Groups): <1 year (sexes combined), 1 to 3 years (sexes combined), 4 to 8 years (sexes combined), and males and females separately for ages 9 to 13 years, 14 to 18 years, 19 to 30 years, 31 to 50 years, 51 to 70 years, and 71 years or above, a minimum of 80 respondents in each DRI was assigned to each province and the remainder were allocated using a power allocation technique (48). In addition, the provincial governments of Ontario, Manitoba, and Prince Edward Island (P.E.I.) paid for their provincial sample buy-ins. The target population did not encompass full-time members of the Canadian Forces or people living in the Territories, on First Nation Reserves or Crown Lands, in prisons or care facilities, or in some remote areas. The high response rate, accompanied with statistical adjustment for non-response, forms the basis to use the results of this survey as representative of the population (refer to Table 3.1).

Table 3.1: Sample Size and Response Rate for the CCHS 2.2 by Province and for Canada (49)

Province	Actual Sample	Response Rate (%)
Newfoundland and Labrador	1,734	83.3
Prince Edward Island	1,430	79.2
Nova Scotia	1,705	78.6
New Brunswick	1,633	75.7
Quebec	4,780	75.8
Ontario	10,921	72.7
Manitoba	4,194	82.7
Saskatchewan	2,041	77.1
Alberta	3,021	77.4
British Columbia	3,648	77.1
CANADA	35,107	76.5

Survey Components

CCHS 2.2 encompasses several aspects of Canadians' Health as evident from the questionnaire of the survey. The modules of the survey are as follows: (1) Household and education (all ages) (2) 24-hr recall (all ages) (3) General health (age $\geq$12 years) (4) Physical activities (age $\geq$ 12 years) (5) Sedentary activities (age 12 to 17 years) (6) Children's physical activity (age 6 to 11 years) (7) Self-reported height and weight (age $\geq$ 18 years) (8) Vitamin and mineral supplements (all ages) (9) Vitamin and mineral supplement details (all ages) (10) Measured height and weight (age $\geq$ 2 years) (11) Women's health (age $\geq$ 9 years) (12) Fruit and vegetable consumption (age $\geq$ 6 months)* (13) Chronic conditions (all ages)* (14) Smoking (age $\geq$ 12 years) (15) Alcohol (age $\geq$ 12 years) (16) Food Security (all households) (17) Sociodemographic characteristics (all ages)* (18) Labour force participation (age 15 to 75 years) (19)

Income (all ages) (20) Administration (data sharing) (all ages). The components marked with asterisk are contributing survey modules for this research study.

3.3 Methodology/Approach

The notion of simple random sampling (SRS) that all sample observations are independently selected with equal probabilities of selection is rarely met for complex survey data where some sample observations may be weighted more heavily than others, and some are included in the sample by virtue of their membership in a certain group rather than being selected independently. As a result, the analysis of complex social surveys using simple random sampling with replacement (SRSWR) assumption turns out to be biased and misleading. Any survey that imposes restrictions on the sampling beyond those of SRSWR is complex in design and requires special analytic considerations. Lohr has succinctly described the two basic approaches that have widely been used in regression analysis (50):

(1) *Design-based:* Inferences are based on repeated sampling from the finite population, and the probability structure used in inference is that defined by the random variables indicating inclusion in the sample. A model that generates the data may exist, but it is not necessarily known what it is, so the analysis does not rely on any theoretical model. Weights are needed for estimating population characteristics and by analogy should be used in linear regression as well.

(2) *Model-based:* A stochastic model describes the relation between y_i and x_i that holds for every observation in the population. One possible model is

$Y_i | x_i = x_i^T \beta + \varepsilon_i$, with the ε_i's independent and normally distributed with constant variance. If the observations in the population really follow the model, then the sample design should have no effect as long as the probabilities of selection depend on y only through the x's.

As CCHS is complex in design, design-based approach is applied to account for the issues of stratification and clustering. The CCHS (Cycle 2.2) primarily used the area frame designed for the Canadian Labour Force Survey (LFS) to select the sample of households. The sampling plan of the LFS is a *multistage stratified cluster design* in which the dwelling is the final sampling unit. Variance estimation of a complex design is another important part of design-based approach. The most popular methods used for variance estimation are:

(1) Linearization Technique

(2) Taylor Expansion Method

(3) Replication Methods (Balanced Repeated Replication, Jackknife, Bootstrap)

As per guideline of Statistics Canada, a bootstrap technique has been employed to approximate variance of estimators of CCHS 2.2 through BOOTVAR macro, a SAS based program developed by Statistics Canada and intended for use with bootstrap weights distributed with CCHS 2.2 to estimate totals, ratios/proportions, difference between two ratios/proportions, percentiles, linear and logistic regressions, chisquare and their variances.

Bootstrapping Complex Sampling Designs

The seminal paper of applying the bootstrap resampling method to stratified multistage design was developed by Rao and Wu (51) and extended again by Rao, Wu and Yue (52). This latter design was implemented in the CCHS. It assumes L design strata, where stratum h has N_h clusters and $n_h \geq 2$ sampled clusters. Subsampling within selected clusters is performed according to some probability sampling design with unbiased estimation of cluster totals, Y_{hi}, with $h = 1,\ldots,L$ and $i = 1,\ldots,n_h$.

An estimator of the total Y, for example, is obtained using the variable of interest, y_{hik}, and design weights, w_{hik}, associated with the kth sample element in sample cluster i belonging to stratum h by

$$\hat{Y} = \sum_{(hik)\in s} w_{hik} y_{hik},$$

where s denotes the sampled elements. The CCHS design weights are then poststratified to ensure consistency with known demographic totals. Given that each element in the population belongs to a poststratum that can cut across the design strata, the total number of elements in the c-th poststratum is ${}_{c}M$, a known quantity. Letting ${}_{c}w_{hik}$ represent the poststratified or final weight defined by

$${}_{c}w_{hik} = \frac{{}_{c}M}{{}_{c}\hat{M}} w_{hik},$$

${}_c\hat{M} = \sum_{(hik)\in s} w_{hik}\ {}_c\delta_{hik}$ and ${}_c\delta_{hik}$ the poststratum indicator variable, the poststratified estimator is defined as

$$\hat{Y}_{ps} = \sum_{c} \sum_{(hik)\in s} {}^{c}w_{hik}\, y_{hik}\ {}_c\delta_{hik}\ .$$

The standard bootstrap variance estimator for $\hat{\theta} = g(\hat{Y})$, is calculated as follows:

(i) Independently for each stratum, select a simple random sample of $n_h - 1$ clusters with replacement from the n_h sample clusters.

(ii) Fore each resample $r\ (r = 1,2,\ldots,R)$

$$w_i(r) = w_i \times \frac{n_h}{n_h - 1} m_i(r)$$

where $m_i(r)$ is the number of times that observation i is selected to

be

in the resample. Calculate $\hat{\theta}^*_r$, using the weights $w_i(r)$

(iii) Repeat steps (i) and (ii) R times, for R a large number.

(iv) Calculate

$$\hat{V}_B(\hat{\theta}) = \frac{1}{R-1} \sum_{r=1}^{R} \left(\hat{\theta}^*_r - \hat{\theta}\right)^2$$

Bootstrap technique has been applied in all estimation techniques including estimation of racial distribution, estimation of proportion of less than 5 times/servings of fruit and vegetables per day, estimation of differences of proportions of 5 or more times/servings of fruit and vegetables per day and estimation of odds ratio in logistic regression for each ethnic group compared to White ethnic group (reference group) while consuming 5 or more FVI.

3.4 Technical Details on Modelling in Logistic Regression

Model Building Strategy

(i) Fitting univarate models with main effects separately at 0.05 level of significance

Separate univariate logistic regressions are run with each main effect as regressor to assess its significance in determining the probability of having long-term diseases. The most clinically relevant regressors *age* and *sex* and the main covariates *fruit and vegetable consumption* and *race/ethnicity* were found significant to be included in the model.

(ii) Checking linearity in the logit for continuous variables

The only continuous variable *age* was checked for the linearity in the logit assumption by examining smoothed scatter plots of the percent of subjects diagnosed with long-term diseases vs the continuous variable *age*. The scatter plot justifies the inclusion of *age* as linear effect as percentage of long-term diseased people over different age segments exhibit nearly a smoothed linear line.

(iii) Checking for interactions of main effects

After finding an appropriate scale (i.e. linear effect) for the continuous variable, the next step was to check the possible interaction among the main effects. Model interaction terms were judged against twice the positive deviance between the log-likelihood of the full model (with interaction terms) and of the reduced model (without interaction terms i.e. only the main effects). The interaction between *fruit and vegetables* & *race/ethnicity* was found insignificant at 0.05 level of significance.

(iv) Applying Forward Stepwise Selection procedure to finalize the best model

To select the final multivariable model, stepwise regression was performed with Forward Selection procedure at 0.05 level of significance. *Ages, sex, fruit and vegetable consumption, race/ethnicity are* all included in the final model for predicting the probability of having long-term diseases.

Model and Variables Used

Logistic regression with main outcome of interest *diseases*, exposure variable *fruit and vegetable consumption*, confounder *age, sex* and classifier predictor *racial/cultural origin* is applied here to justify the conjectures (i) whether people who have taken five or more servings of fruits and vegetables exhibit less chance of developing *diseases* compared to those who haven't; (ii) whether some *racial/cultural groups* have more likelihood of having diseases compared to other *racial/cultural groups*; and (iii) how the confounders *age* and *sex* act in the causal pathway between *diseases* and *fruit and vegetable consumption*. The linear logistic model (Koch and Edwards (53)) is given by

$$\text{logit}\, P(X) = \alpha + \beta E + \sum_{i=1}^{p_1} \gamma_i V_i + \sum_{j=1}^{p_2} \delta_j C_j$$

where α is intercept term, E indicates exposure (*fruit and vegetable consumption*) variable, V 's (*age, sex*) are thought to account for confounding in the data, C 's (*racial* *groups*) are classifier predictors. The coefficients β and δ represent the change in the log odds that would result from a one unit change in the respective variable when other variables are fixed. Estimates of the model parameters and their corresponding odds ratio, standard errors, Wald statistic, p-value, 95% confidence intervals for the odds ratio are produced with BOOTVAR macro that accurately reflects the complex sample design.

Outcome of Interest

The outcome variable of interest is a two-level (binary) variable that takes on the value *1* if the respondent's "long-term conditions" are expected to last or have already lasted 6 months or more and that have been diagnosed by a health professional and *0* if the respondent does not have that attribute. As part of the several logistic regression runs, the exhaustive listing of the all outcomes of interest are as follows:

(1) *blood_pressure* = 1: if the respondent has high blood pressure

0: if the respondent does not have high blood pressure

(2) *diabetes* = 1: if the respondent has diabetes

0: if the respondent does not have diabetes

(3) *heart_disease* = 1: if the respondent has heart disease

0: if the respondent does not have heart disease

(4) *cancer* = 1: if the respondent has cancer

0: if the respondent does not have cancer

(5) *intestinal_ulcers* = 1: if the respondent has intestinal or stomach ulcers

0: if the respondent doe not have intestinal or stomach ulcers

(6) *bowel_disorder* = 1: if the respondent has bowel disorder such as Crohn's Disease or Colitis

0: if the respondent does not have bowel disorder

(7) *osteoporosis* = 1: if the respondent has osteoporosis

0: if the respondent does not have osteoporosis

(8) *long_term* = 1: if the respondent has any other long-term physical or mental health condition that has been diagnosed by a health professional

0: if the respondent does not have any other long-term physical or mental health condition

(9) *chronic* = 1: if the respondent has chronic health conditions

0: if the respondent does not have chronic health conditions

(It is a derived variable based on variables from 1 to 8)

Main Covariate of Interest

(1) While covariate involves Total Fruit and Vegetables, the predictor is a dichotomous variable complying with the Canadian Food Guide threshold:

_5ormoreserv = 1: if the respondent takes 5 or more times/servings of fruits and vegetables per day

0: if the respondent takes less than 5 times/servings of fruits and vegetables per day

It is strongly felt that the association between *diseases* and constituent parts of Total Fruit and Vegetables *viz.* Fruit Juice, Fruit, Green Salad, Potatoes, Carrots, Other Vegetables would provide a deeper insight of the interplay between *diseases* and *fruit and vegetable consumption*. Hence for other logistic regression analyses, the following covariates, measured on a continuum (times per day), are incorporated in several runs:

(a) FVCDDJUI (How often do you drink fruit juice per day?)

(b) FVCDDFRU (Not counting juice, how often do you usually eat fruit per day?)

(c) FVCDDSAL (How often do you usually eat green salad per day?)

(d) FVCDDPOT (How often do you usually eat potatoes, not including French fries, fried potatoes or potato chips, per day?)

(e) FVCDDCAR (How often do you usually eat carrots per day?)

(f) FVCDDVEG (Not counting carrots, potatoes, or salad, how many servings of other vegetables do you usually eat per day?)

(2) Another major covariate of analysis is *Cultural/Racial Origin* which is tapped by SDCDDRAC variable that contains 14 categories: *White, Black, Korean, Filipino, Japanese, Chinese, Aboriginal People of North America, South Asian, Southeast Asian, Arab, West Asian, Latin American, Other Racial or Cultural Origin, Multiple Racial or Cultural Origin*. The *Multiple Racial or Cultural Origin* has been dropped from the analysis as it poses the problem of indeterminate mixing of cultural groups. The convention of incorporating a nominal exposure variable that has *k* categories in logistic regression model is using *k-1* categories that will reflect the nominal exposure variable leaving the *reference* category (51). SDCDDRAC is decomposed into 13 dichotomous

variables and 12 classifier predictors are selected for analysis leaving *White* as *reference* group. Again, due to low volume of cell frequency in some racial groups per disease which is evident from cross-tabulations of *diseases* by *cultural/racial groups*, *Korean, Filipino, Japanese, Chinese* are amalgamated into a group namely *KFJC* and *South Asian, Southeast Asian, Arab, West Asian* origins are merged into a new cultural group named *Asian*, to meet the strict disclosure policy of Statistics Canada. And the last but not least, although the final classifier predictors are *Black, KFJC, AbNA (Aboriginal People of North America), Asian, Latin American, Other*, some racial groups are dropped from the analysis when they have *less than five cell frequency* for any specific *disease.*

Potential Confounders considered in the Model

DHHD_AGE and *DHHD_SEX* are two confounder variables considered in all logistic regression runs and presumed to have much more importance as potential confounders in many nutrition and health survey analyses. *DHHD_AGE* is a continuous variable and *DHHD_SEX* is a dichotomous variable indicating two labels: 1 if Male, 0 if Female.

Goodness of Fit Test for Logistic Model with Sparse Data

There are two different approaches to assessing goodness of fit test in logistic regression model. The first, known as residual analysis, examines the model which on the level of individual observations and looks for those observations which are not adequately described by the model or which are highly influential on the model fit. The second approach to goodness of fit test seeks to combine the information on the amount of lack-of-fit in a single number. Statistical tests, so called goodness-of-fit tests, are then performed to judge if the lack of fit is statistically significant or due to random chance. These tests, however, have serious limitations with sparse data, where sparse data means that for every pattern of covariate values there exist only small number of observations (54, 55). CCHS 2.2 with long-term diseases as outcome variable is thinly distributed (sparse) with respect to each covariate pattern. While long-term disease conditions per se are less prevalent, many ethnic groups do not carry many of the long-term conditions or in some cases exhibit very low frequency.

To assess the goodness of fit test for logistic regression one in general calculates the Pearson statistic

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - m_i\hat{\pi}_i)^2}{m_i\hat{\pi}_i(1-\hat{\pi}_i)}$$

or the Deviance statistic

$$D = 2\sum_{i=1}^{N} y_i \log\left(\frac{y_i}{m_i\hat{\pi}_i}\right) + (m_i - y_i)\log\left(\frac{m_i - y_i}{m_i(1-\hat{\pi}_i)}\right)$$

Both rely on the premise of comparing observed y_i to predicted $m_i\hat{\pi}_i$ and their statistical significance is judged comparing to a χ^2_{N-p-1} distribution. The validity of this distribution, however, relies on the assumption of large m_i which is unrealistic with continuous covariates and both tests show unsatisfactory behaviour with sparse data, that is, small m_i. To circumvent this problem, an alternative goodness-of- fit-test, the so called *Hosmer-Lemeshow* (56) test was introduced which relies on a new grouping of the individual observations in approximately ten groups with roughly the same size where the grouping depends on the percentiles of the estimated probabilities ($\hat{\pi}_i$) from the model. Observed and expected numbers of events are determined for each of the new groups, and their discrepancies are summed. Lack-of-fit is judged by comparing the sum, which is, after standardization, a Pearson statistic from a $2 \times g$ table with g being the number of new groups, to a χ^2_{g-2} distribution. This test has some deficiencies however. The value of the test statistic might depend on the number of new groups and on the calculating algorithm. However, there are additional tests procedures that do not rely on the assumption of large cell counts and outperform Hosmer-Lemeshow-goodness-of-fit-test (55). Osius and Rojek (57) proposed a statistical test by standardizing χ^2 with derived asymptotic moments and comparing the resulting test statistic to the standard normal distribution. *McCullagh* (58) made a similar proposal and relaxed the assumption of large m_i but he argued to use conditional asymptotic moments for χ^2 given the parameter estimates. His test statistic is also compared, after standardization with the conditional asymptotic moments, to a standard normal distribution.

%GOFLOGIT, a SAS/IML macro, has been used to carry out the Standard Pearson, Standard Deviance, Osius-Test, and Mc-Cullagh Test on the estimated logistic model to assess its goodness of fit. As all logistic regression runs involve total fruit and vegetables and its constituent parts at several stages with all other variables remaining the same, goodness of fit tests of the logistic models involving total fruit and vegetables would be a good indicator of the overall fit.

Chapter Four:

Empirical Analysis and Results

4.1 Introduction

This chapter is intended to map out the empirical analysis and results of this ethnicity-based research. *Section 4.2* deals with distribution of cultural/racial origin in Canada expressing the perspective that Canadian population's differential cultural/racial distribution provides the foundation for the data-analytic focus of this study. *Section 4.3* describes the percentage of each ethnic groups lying above or below the 5 times/servings of fruits and vegetables per day benchmark, their mean frequency consumption of fruit and vegetables, and finally takes up the differences among fruit and vegetable consumption of all ethnic groups which would manifest that long-term diseases ensue when ethnic groups have substantially low volume of servings of fruit and vegetables. *Section 4.4* sums up the results of logistic regression analyses that use the bootstrap estimation technique. Several goodness of fit tests of the estimated logistic models are presented in *Section 4.5*.

4.2 Canadian Ethnic Mosaic

Ethnicity-based research lays the claim that a racial/cultural group refers to a population who share common cultural characteristics such as language, religion, and diet and assumes that important genetic/cultural differences exist among cultural groups (10). Thus the determinant of cultural mosaic is recognized as *Ethnic Identity* which relies more on a shared cultural definition of identity than solely on biological similarity which is manifested through *Ethnic Ancestry*. Given that variations in disease rates between populations may be explained by socioeconomic, socio-cultural, biological, and genetic factors, classification by cultural origin rather than ethnic origin is desirable (10). Canadian cultural diversity has13 independent categories (White, Black, Korean, Filipino, Japanese, Chinese, Aboriginal People of North America, South Asian, Southeast Asian, Arab, West Asian, Latin American, Other Racial/Cultural Origin) and one multiracial category. Table 4.1 shows that *White* appears to be the most dominant cultural group accounting for 82.40% of the whole population while the other ethnic groups have relatively low percentages. The second largest group includes *South Asian* which constituted 3.74%, *Chinese* (3.14%), and *Black* (2.05%). While multiculturalism is a defining characteristic of Canada's contemporary national identity, the reality of *Multicultural/Multiracial Origin* is both complex and problematic as this is an eclectic group made up of several minorities which fails to represent any meaningful group to compare others to, although *Multicultural Origin* makes up 1.50% of the total population. *Aboriginal People of North America*, *Other*, *Filipino*, and *Southeast Asian* groups make up 1.28%, 1.27%, 1.24% and 1.01% respectively, whereas *Arab* (0.89%), *Latin American* (0.68%), *West Asian* (0.38%), *Korean* (0.28%), *Japanese* (0.15%) belong to the least-contributing group. The percentages of male and female in all ethnic groups are fairly close to each other, with female (50.64%) proportion exceeding slightly that of male (49.36%) on average.

<u>Table 4.1:</u> Estimated Distribution of Cultural/Racial Origins by Sex

	Male		Female		Total	
	Weighted Frequency	Per cent	Weighted Frequency	Per cent	Weighted Frequency	Per cent
White	12,520,000 (approx.)	40.41	13,010,000 (approx.)	42.00	25,530,790	82.40
Black	346,990	1.12	287,489	0.93	634,479	2.05
Korean	37,424	0.12	48,767	0.16	86,190.48	0.28
Filipino	154,982	0.50	229,257	0.74	384,238.7	1.24
Japanese	17,377	0.06	27,638	0.09	45,014.99	0.15
Chinese	560,033	1.81	412,374	1.33	972,406.1	3.14
Aboriginal People of North America	173,753	0.56	221,888	0.72	395,640.9	1.28
South Asian	622,910	2.01	534,335	1.72	1,157,245	3.74
Southeast Asian	134,586	0.43	177,694	0.57	312,279.9	1.01
Arab	184,187	0.59	90,729	0.29	274,915.7	0.89
West Asian	68,479	0.22	50,546	0.16	119,025	0.38
Latin American	97,359	0.31	113,409	0.37	210,768.7	0.68
Other Racial/Cultural Origin	160,253	0.52	232,698	0.75	392,950.8	1.27
Multiple Racial/Cultural Origin	216,370	0.70	249,792	0.81	466,162.2	1.50
Total	15,290,000 (approx.)	49.36	15,690,000 (approx.)	50.64	30,982,107	100.00

4.3 A Juxtaposition of Cultural/Racial Groups' Fruit and Vegetable Consumption

Food has long been regarded as a useful ethnic marker, a way of defining who we are. In the spirit of Canadian Food Guidelines, proportion of cultural/racial groups consuming less than 5 times/servings of fruit and vegetables (FV) per day and 5 or more times/servings of FV per day are estimated using bootstrap technique as CCHS 2.2 is a complex survey. Table 4.2 displays that Southeast Asian (13%), Aboriginal People of North America (15%), West Asian (17%), Korean (18%), Japanese (19%), Chinese (23%) people have a very low percentages who consume 5 times/servings or more of FV per day. An estimated 65 per cent of White people, the major ethnic group, fail to eat adequate number of fruit and vegetables servings per day. The next dominant ethnic group, South Asian people, also exhibits that only 28 per cent of them are able to have 5 or more times/servings of FV per day. Filipinos (69%) and Blacks (63%) also possess considerably a high proportion not eating 5 or more servings of FV per day. In contrast, a relatively higher percentage of Arabs, Latin Americans and Other racial/cultural groups, nearly 40 per cent, maintain 5 or more times/servings of FV per day. In summary, the majority of all ethnic groups in Canada are lying below the '5 or more times/servings of FV per day' benchmark which might portrait an important aspect of nutritional status of Canadian diverse population.

Table 4.2: Proportion of Cultural/Racial Groups consuming 5 or more times/servings of Total Fruit and Vegetables per day

		White	Black	Korean	Filipino	Japanese	Chinese	Aboriginal People of North America	South Asian	South-east Asian	Arab	West Asian	Latin American	Other
5 or More times/servings per day	**Proportion**	**0.3457**	**0.3602**	**0.1745**	**0.3054**	**0.1824**	**0.2292**	**0.1494**	**0.2866**	**0.1332**	**0.3887**	**0.1729**	**0.3899**	**0.4093**
	S.E.	0.0056	0.0422	0.0712	0.0595	0.0825	0.0265	0.0220	0.0336	0.0394	0.0839	0.0888	0.0888	0.0590
	CV	1.61	11.71	40.81	19.49	45.24	11.55	14.73	11.71	29.58	21.57	51.37	22.78	14.42
	95%LCL	0.3348	0.2775	0.0349	0.1887	0.0207	0.1773	0.1063	0.2209	0.0560	0.2244	-0.0012	0.2158	0.2936
	95%UCL	0.3566	0.4429	0.3141	0.4221	0.3442	0.2811	0.1925	0.3524	0.2105	0.5531	0.3470	0.5640	0.5250

It is worth noting that Statistics Canada has a strict disclosure policy and sampling variability guidelines for CCHS survey estimates (Table 4.3):

Table 4.3: Statistics Canada Sampling Variability Guidelines for CCHS

Type of Estimate	CV	Guidelines
Acceptable	0.0 -16.5	General unrestricted release
Marginal	16.6 – 33.3	General unrestricted release but with warning cautioning users of the high sampling variability. Should be identified by letter M
Unacceptable	> 33.3	No release. Should be flagged with letter U.

Estimated percentages of all ethnic groups consuming less than 5 times/servings of FV per day have Coefficient of Variation (CV) within the acceptable limit of Statistics Canada sampling variability guidelines, whereas some racial groups, as for instance, Korean, Japanese and West Asian exhibit unacceptable CV while estimating percentages consuming 5 or more times/servings of FV per day. Therefore, their results should be viewed with caution as standard errors are large due to their small numbers in that category. Mean frequency consumption of total fruit and vegetables and its constituent components viz., fruit juice, fruit, salad, potato, carrot, other Vegetables are also estimated to have a more in-depth look into the synthesis of which ethnic groups reveal significantly lower consumption of FV per day (refer to Table 4.4). It has been observed that Korean, Japanese, Chinese, Aboriginal People of North America, Southeast Asian and West Asian people's mean intake of FV per day is lower compared to other ethnic groups, which obviously bolsters the findings elicited from Table 4.2.

Table 4.4: Mean Frequency Consumption of Fruit and Vegetables for Cultural/Racial Groups

	Fruit Juice	Fruit	Salad	Potato	Carrot	Other Vegetables	Total Fruit and Vegetables
White	0.90 (0.01)	1.21 (0.01)	0.44 (0.01)	0.37 (0.01)	0.36 (0.01)	1.07 (0.01)	4.37 (0.02)
Black	1.21 (0.08)	1.05 (0.08)	0.42 (0.03)	0.20 (0.02)	0.40 (0.04)	0.85 (0.05)	4.16 (0.19)
Korean	0.89 (0.19)	0.85 (0.10)	0.49 (0.08)	0.18 (0.04)	0.25 (0.04)	1.08 (0.09)	3.79 (0.28)
Filipino	1.05 (0.12)	1.33 (0.12)	0.32 (0.04)	0.13 (0.02)	0.31 (0.03)	1.06 (0.11)	4.23 (0.25)
Japanese	0.81 (0.18)	0.67 (0.23)	0.36 (0.07)	0.23 (0.03)	0.30 (0.07)	1.33 (0.17)	3.71 (0.44)
Chinese	0.57 (0.04)	1.15 (0.05)	0.25 (0.02)	0.13 (0.01)	0.25 (0.02)	1.40 (0.05)	3.78 (0.10)
Aboriginal People of North America	0.84 (0.05)	0.81 (0.05)	0.32 (0.02)	0.46 (0.03)	0.29 (0.02)	0.76 (0.05)	3.50 (0.11)
South Asian	0.81 (0.06)	1.12 (0.06)	0.44 (0.03)	0.30 (0.02)	0.37 (0.02)	1.15 (0.07)	4.24 (0.15)
Southeast Asian	0.66 (0.07)	1.09 (0.12)	0.38 (0.06)	0.15 (0.03)	0.23 (0.04)	0.94 (0.08)	3.49 (0.19)
Arab	0.99 (0.18)	1.74 (0.25)	0.73 (0.13)	0.26 (0.06)	0.27 (0.07)	0.93 (0.14)	4.98 (0.57)
West Asian	0.90 (0.20)	0.92 (0.20)	0.65 (0.08)	0.21 (0.04)	0.21 (0.05)	0.76 (0.10)	3.68 (0.45)
Latin American	1.26 (0.23)	1.05 (0.09)	0.51 (0.06)	0.35 (0.07)	0.41 (0.08)	0.71 (0.07)	4.36 (0.30)
Other	1.02 (0.11)	1.27 (0.08)	0.54 (0.05)	0.32 (0.05)	0.39 (0.04)	1.07 (0.10)	4.64 (0.23)

* Values in parentheses indicate standard errors as results generated using 500 bootstrap replicates

As part of dispelling the myths about group disparities, differences in proportions of ethnic groups consuming 5 or more times/servings of FV per day are estimated using BOOTVAR macro of SAS. Although substantive differences among ethnic groups are not seemingly evident from the table of mean frequency consumption, statistically significant differences exist among these groups (refer to Table 4.5) which would prop up the earlier assertions. Aboriginal People of North America is having significant differences with 8 ethnic groups except Korean, Japanese, Southeast Asian and West Asian people in eating 5 or more times/servings of FV per day. Southeast Asian is another highest differenced group in the comparison table who does not have intake differences with Korean, Japanese, Aboriginal People of North America and West Asian. It might be argued that as Korean, Japanese, Southeast Asian, Aboriginal People of North America, and West Asian have least consumption of 5 or more times/servings of fruit and vegetables (refer to Table 4.2), the deviations among them are insignificant and hence Aboriginal People of North America, Southeast Asian do not demonstrate wide gap with the aforementioned groups. The undefined or assorted Other ethnic group, having the highest intake of 5 or more times/servings of FV per day among all ethnic groups, displays potential gap in consumption of FV per day with six ethnic groups. Only 22 per cent of Chinese people in Canada take more than 5 times/servings of fruit and vegetables per day and shows significant difference with that of White, Black, Aboriginal People of North America, Southeast Asian and Other groups.

Table 4.5: Estimated Difference in Proportions (p-value) of Racial/Cultural Groups Consuming 5 or More Times/Servings of Fruits and Vegetables per Day (using 500 bootstrap replicates)

	White	Black	Korean	Filipino	Japanese	Chinese	Aboriginal People of North America	South Asian	Southeast Asian	Arab	West Asian	Latin American
White												
Black	-0.01 (0.73)											
Korean	**0.17 (0.02)**	**0.19 (0.02)**										
Filipino	0.04 (0.50)	0.05 (0.45)	-0.13 (0.13)									
Japanese	**0.16 (0.05)**	0.18 (0.06)	-0.01 (0.94)	0.12 (0.23)								
Chinese	**0.12 (<0.01)**	**0.13 (<0.01)**	-0.05 (0.46)	0.08 (0.24)	-0.05 (0.59)							
Aboriginal People of North America	**0.20 (<0.01)**	**0.21 (<0.01)**	0.03 (0.73)	**0.16 (0.01)**	0.03 (0.69)	**0.08 (0.02)**						
South Asian	0.06 (0.08)	0.07 (0.19)	-0.11 (0.15)	0.02 (0.78)	-0.10 (0.25)	-0.06 (0.19)	**-0.14 (<0.01)**					
Southeast Asian	**0.21 (<0.01)**	**0.23 (<0.01)**	0.04 (0.60)	**0.17 (0.02)**	0.05 (0.60)	**0.10 (0.04)**	0.02 (0.72)	**0.15 (<0.01)**				
Arab	-0.04 (0.61)	-0.03 (0.76)	-0.21 (0.06)	-0.08 (0.41)	-0.21 (0.06)	-0.16 (0.07)	**-0.24 (<0.01)**	-0.10 (0.24)	**-0.26 (<0.01)**			
West Asian	**0.17 (0.05)**	0.19 (0.06)	0.01 (0.99)	0.13 (0.22)	0.01 (0.94)	0.06 (0.55)	-0.02 (0.80)	0.11 (0.23)	-0.04 (0.68)	0.22 (0.07)		
Latin American	-0.04 (0.62)	-0.03 (0.77)	-0.22 (0.06)	-0.08 (0.45)	-0.21 (0.08)	-0.16 (0.09)	**-0.24 (<0.01)**	-0.10 (0.29)	**-0.26 (<0.01)**	-0.01 (0.99)	-0.22 (0.08)	
Other	-0.06 (0.28)	-0.05 (0.48)	**-0.23 (0.01)**	-0.10 (0.23)	**-0.23 (0.03)**	**-0.18 (<0.01)**	**-0.26 (<0.01)**	-0.12 (0.07)	**-0.28 (<0.01)**	-0.02 (0.83)	**-0.24 (0.03)**	-0.02 (0.85)

4.4 Estimation of Odds Ratios of developing Long-term Diseases

Fruit and vegetables vs. long-term diseases

Table 4.6 reflects that consumers of 5 or more times/servings of FV per day show less likelihood to have diabetes (odds ratio: 0.76, p value: 0.0240), long-term health conditions (odds ratio: 0.89, p value: 0.0305), and chronic health conditions (odds ratio: 0.86, p value: 0.0122) than those who take less than 5 times/servings of FV per day. Logistic regression performed on each component of fruit and vegetables instead of total fruit and vegetables gives more convincing results. Those who drink more Fruit juice (odds ratio: 0.71, p value: 0.0017) have demonstrated less propensity to develop diabetes (refer to Table 4.7). There is compelling evidence from Table 4.8 that eating plenty of fruits (not counting fruit juice) can help Canadians ward off heart diseases (odds ratio: 0.90, p value: 0.0476), long-term disease conditions (odds ratio: 0.93, p value: 0.0026) and chronic disease conditions (odds ratio: 0.93, p value: 0.0018). This evidence supports the findings of the study of Hung and Joshipura that the higher the average daily intake of fruits, the lower the chances of developing cardiovascular diseases (5). Green Salad is also an important source of nutrients and provides guard against heart disease (odds ratio: 0.68, p value: 0.0016), intestinal ulcers (odds ratio: 0.56, p value: 0.0078), long-term conditions (odds ratio: 0.83, p value: 0.0035) and chronic conditions (odds ratio: 0.74, p value< 0.01) (refer to Table 4.9). More interestingly, Potatoes (odds ratio: 1.56, p value: 0.0300) have found to be associated in exacerbating cancer disease as evident from Table 4.10. Although there is no significant association observed between carrot and heart disease & cancer, carrot appears to be effective in attenuating the risk of intestinal ulcers (odds ratio: 0.42, p value: 0.0015), bowel disorders (odds ratio: 0.60, p value: 0.0152) and long-term conditions (odds ratio:0.84, p value:0.0332) (refer to Table 4.11). Other vegetables (not counting carrots, potatoes or green salads) also indicate that its high consumption would protect Canadian from developing diabetes (odds ratio: 0.86, p value: 0.0370), intestinal ulcers (odds ratio: 0.81, p value: 0.0278), bowel disorder (odds ratio: 0.60, p value: 0.0152), and long-term disease conditions (odds ratio: 0.94, p value: 0.0378). Lower consumption of any component of fruit and vegetables or total fruit and vegetables does not demonstrate any apparent predisposition to the development of high blood pressure or osteoporosis.

Ethnicity vs. long-term diseases

The burden of long-term diseases falls disproportionately on racial/cultural groups in Canada. Several long-term diseases have been found to be prevalent in Aboriginal people of North America. This ethnic group has an increased chance of developing high blood pressure (odds ratio: 1.86, p value: 0.0126), diabetes (odds ratio: 3.61, p value: 0.0001), heart disease (odds ratio: 1.98, p value: 0.0342), intestinal ulcers (odds ratio: 2.34, p value: 0.0024), chronic conditions (odds ratio: 1.42, p value: 0.0103) compared to White ethnic group. Aboriginal People of North Americans high susceptibility to the significant long-term diseases may be ascribable to lower intake of fruit and vegetables (refer to Table 4.2). The Black, KFJC, Asian, Latin American and Other ethnic groups (compared to White *reference group*) do not appear to have significant likelihood to develop high blood pressure and diabetes. In contrast, the participants of the Black, KFJC and Asian have reported themselves more protected against long-terms disease conditions and chronic conditions. The odds ratio of developing long-term disease conditions for the Black, KFJC and Asian ethnic people are 0.61 (p value: 0.0201), 0.48 (p value: 0.0002), 0.64 (p value: 0.0134) respectively and the odds ratio of developing chronic conditions for the KFJC and Asians are 0.71 (p value: 0.0208) and 0.71 (p value: 0.0274) respectively.

4.5 Goodness of Fit Test for the Estimated Logistic Model

As CCHS 2.2 is very sparsely distributed with few frequencies of long-term diseases (successes) for each covariate pattern, the traditional Hosmer-Lemeshow goodness of fit test is outperformed by Osius-Rojek and McCullagh goodness of fit test (refer to Table 4.13). While modelling the probability of having blood pressure, Pearson (p-value: 0.01), Deviance (p-value: 0.02), and Hosmer-Lemeshow (p-value: 0.01) tests do not support the appropriateness of the model but the Osius-Rojek (p-value: 0.10) and McCullagh (p-value: 0.08) tests are relatively better in validating its appropriateness. In modelling diabetes, Deviance (p-value: 0.99), Osius-Rojek (p-value: 0.23), McCullagh (p-value: 0.21) outperform the Person (p-value: 0.05) and Homser-Lemeshow (p-value: 0.00). Only the Deviance (p-value: 0.55, p-value: 0.60) statistic among all other goodness of fit tests evaluates the prognosis of the heart disease and osteoporosis respectively as a good fit. The goodness of fit of the model with cancer, intestinal ulcers, bowel disorder drastically improves in almost all goodness of fit tests except Hosmer-Lemeshow goodness of fit test. Interestingly, the logistic models with other long-term health conditions and chronic conditions appear to be not-well-fitted models by all goodness of fit criteria. These models contain maximum number of covariates and highest number of covariate patterns. Kuss (2002) (55) depicted these circumstances in words like “In general, all tests gain power with increasing m_i. All in all, there is low power for detecting lack-of-fit with small m_i".

To put it simply, as the number of covariate increases in these models, the total number of covariate pattern increases and hence the data becomes increasingly sparse. Another reasoning might be that some racial groups with different age, sex, and fruit and vegetable consumption may have few or no long-term diseases which leads to the problem of the sparsity.

Table 4.6: Odds Ratios from the Logistic Regression of Long-term Health Conditions and Total Fruit & Vegetable Consumption, Race/Ethnicity

Outcome of Interest		5ormoreserv	Black	KFJC	AbNA	Asian	Latin American	Other
Blood Pressure	OR	0.91	**2.07**	1.15	**1.86**	0.95		0.46
	p-value	0.2044	**0.0575**	0.5479	**0.0126**	0.8420		0.0803
	95%CI for OR	(0.78, 1.05)	**(0.98, 4.38)**	(0.73, 1.90)	**(1.14, 3.02)**	(0.59, 1.55)		(0.19, 1.10)
Diabetes	OR	**0.76**	1.89	1.10	**3.61**	1.26		
	p-value	**0.0240**	0.1831	0.8412	**0.0001**	0.5131		
	95%CI for OR	**(0.59, 0.96)**	(0.74, 4.85)	(0.44, 2.73)	**(1.84, 6.88)**	(0.63, 2.54)		
Heart Disease	OR	0.85			**1.98**			
	p-value	0.1205			**0.0342**			
	95%CI for OR	(0.69, 1.04)			**(1.05, 3.74)**			
Cancer	OR	1.05			1.95			
	p-value	0.8131			0.1714			
	95%CI for OR	(0.71, 155)			(0.75, 5.09)			
Intestinal Ulcers	OR	0.84	1.12		**2.34**			
	p-value	0.3345	0.8860		**0.0027**			
	95%CI for OR	(0.59, 1.20)	(0.23, 5.43)		**(1.34, 4.08)**			
Bowel Disorder	OR	1.07			1.88			
	p-value	0.6605			0.2201			
	95%CI for OR	(0.80, 1.42)			(0.69, 5.17)			
Osteoporosis	OR	1.01		1.06	1.21	0.65		
	p-value	0.8983		0.8821	0.7642	0.4125		
	95%CI for OR	(0.82, 1.25)		(0.50, 2.25)	(0.35, 4.24)	(0.24, 1.81)		
Long-term	OR	**0.89**	**0.61**	**0.48**	1.14	**0.64**	0.69	0.77
	p-value	**0.0305**	**0.0201**	**0.0002**	0.3450	**0.0134**	0.3930	0.4909
	95%CI for OR	**(0.80, 0.99)**	**(0.41, 0.93)**	**(0.32, 0.71)**	(0.87, 1.49)	**(0.46, 0.91)**	(0.29, 1.62)	(0.37, 1.61)
Chronic	OR	**0.86**	0.84	**0.71**	**1.42**	**0.71**	0.59	0.76
	p-value	**0.0122**	0.3876	**0.0208**	**0.0103**	**0.0274**	0.1958	0.3757
	95%CI for OR	**(0.77, 0.97)**	(0.56, 1.25)	**(0.53, 0.95)**	**(1.09, 1.86)**	**(0.52, 0.96)**	(0.27, 1.31)	(0.42, 1.38)

Table 4.7: Odds Ratios from the Logistic Regression of Long-term Health Conditions and Fruit Juice Consumption, Race/Ethnicity

Outcome of Interest		FVCDDJUI	Black	KFJC	AbNA	Asian	Latin American	Other
Blood Pressure	OR	1.00	2.06	1.16	**1.89**	0.96		0.45
	p-value	0.9824	0.0609	0.5173	**0.0089**	0.8601		0.0767
	95%CI for OR	(0.90, 1.10)	(0.97, 4.37)	(0.74, 1.81)	**(1.17, 3.05)**	(0.59, 1.55)		(0.19, 1.09)
Diabetes	OR	**0.71**	2.03	1.07	**3.75**	1.22		
	p-value	**0.0017**	0.1323	0.8859	**0.0000**	0.5783		
	95%CI for OR	**(0.58, 0.88)**	(0.81, 5.13)	(0.44, 2.61)	**(1.99, 7.06)**	(0.61, 2.46)		
Heart Disease	OR	1.14			**2.12**			
	p-value	0.0774			**0.0124**			
	95%CI for OR	(0.99, 1.32)			**(1.18, 3.83)**			
Cancer	OR	1.16			1.90			
	p-value	0.1088			0.1759			
	95%CI for OR	(0.97, 1.39)			(0.75, 4.82)			
Intestinal Ulcers	OR	1.10	1.09		**2.41**			
	p-value	0.4358	0.9169		**0.0017**			
	95%CI for OR	(0.87, 1.39)	(0.22, 5.30)		**(1.39, 4.16)**			
Bowel Disorder	OR	1.19			1.87			
	p-value	0.0528			0.2289			
	95%CI for OR	(1.00, 1.42)			(0.67, 5.18)			
Osteoporosis	OR	1.02		1.06	1.19	0.66		
	p-value	0.8173		0.8728	0.7833	0.4200		
	95%CI for OR	(0.88, 1.18)		(0.50, 2.26)	(0.34, 4.16)	(0.24, 1.82)		
Long-term	OR	0.99	**0.61**	**0.48**	1.16	**0.65**	0.64	0.77
	p-value	0.8509	**0.0196**	**0.0003**	0.2630	**0.0149**	0.3137	0.4736
	95%CI for OR	(0.94, 1.05)	**(0.41, 0.92)**	**(0.32, 0.71)**	(0.89, 1.52)	**(0.46, 0.92)**	(0.27, 1.52)	(0.37, 1.59)
Chronic	OR	1.03	0.83	**0.72**	**1.48**	**0.72**	0.54	0.75
	p-value	0.2884	0.3622	**0.0286**	**0.0037**	**0.0347**	0.1193	0.3445
	95%CI for OR	(0.97, 1.09)	(0.55, 1.24)	**(0.54, 0.97)**	**(1.14, 1.94)**	**(0.53, 0.98)**	(0.25, 1.17)	(0.42, 1.35)

Table 4.8: Odds Ratios from the Logistic Regression of Long-term Health Conditions and Fruit Consumption, Race/Ethnicity

Outcome of Interest		FVCDDFRU	Black	KFJC	AbNA	Asian	Latin American	Other
Blood Pressure	OR	0.98	2.06	1.16	**1.88**	0.96		0.45
	p-value	0.6702	0.0602	0.5194	**0.0101**	0.8687		0.0779
	95%CI for OR	(0.91, 1.06)	(0.97, 4.37)	(0.74, 1.81)	**(1.16, 3.04)**	(0.59, 1.55)		(0.19, 1.09)
Diabetes	OR	1.00	1.87	1.12	**3.90**	1.27		
	p-value	0.9880	0.1907	0.8022	**0.0000**	0.5013		
	95%CI for OR	(0.91, 1.10)	(0.73, 4.76)	(0.46, 2.75)	**(2.08, 7.30)**	(0.63, 2.58)		
Heart Disease	OR	**0.90**			**2.00**			
	p-value	**0.0476**			**0.0254**			
	95%CI for OR	**(0.82, 1.00)**			**(1.09, 3.66)**			
Cancer	OR	0.93			1.79			
	p-value	0.3566			0.2249			
	95%CI for OR	(0.80, 1.08)			(0.70, 4.62)			
Intestinal Ulcers	OR	0.98	1.11		**2.36**			
	p-value	0.8163	0.8962		**0.0025**			
	95%CI for OR	(0.84, 1.15)	(0.23, 5.37)		**(1.35, 4.12)**			
Bowel Disorder	OR	0.97			1.82			
	p-value	0.6699			0.2479			
	95%CI for OR	(0.84, 1.12)			(0.66, 5.01)			
Osteoporosis	OR	1.00		1.06	1.20	0.66		
	p-value	0.9663		0.8736	0.7791	0.4198		
	95%CI for OR	(0.90, 1.11)		(0.50, 2.26)	(0.34, 4.17)	(0.24, 1.82)		
Long-term	OR	**0.93**	**0.61**	**0.48**	1.13	**0.65**	0.64	0.77
	p-value	**0.0026**	**0.0167**	**0.0002**	0.3745	**0.0161**	0.2972	0.4763
	95%CI for OR	**(0.89, 0.97)**	**(0.40, 0.91)**	**(0.32, 0.71)**	(0.86, 1.47)	**(0.46, 0.92)**	(0.27, 1.49)	(0.37, 1.58)
Chronic	OR	**0.93**	0.83	**0.72**	**1.43**	**0.72**	0.54	0.76
	p-value	**0.0018**	0.3509	**0.0235**	**0.0090**	**0.0355**	0.1246	0.3581
	95%CI for OR	**(0.88, 0.97)**	(0.55, 1.23)	**(0.54, 0.96)**	**(1.09, 1.87)**	**(0.53, 0.98)**	(0.24, 1.19)	(0.42, 1.36)

Table 4.9: Odds Ratios from the Logistic Regression of Long-term Health Conditions and Green Salad Consumption, Race/Ethnicity

Outcome of Interest		FVCDDSAL	Black	KFJC	AbNA	Asian	Latin American	Other
Blood Pressure	OR	0.86	**2.07**	1.13	**1.84**	0.96		0.46
	p-value	0.0970	**0.0590**	0.5965	**0.0140**	0.8780		0.0835
	95%CI for OR	(0.72, 1.03)	**(0.97, 4.40)**	(0.72, 1.77)	**(1.13, 2.99)**	(0.59, 1.57)		(0.19, 1.11)
Diabetes	OR	0.84	1.87	1.09	**3.77**	1.28		
	p-value	0.1912	0.1947	0.8494	**0.0000**	0.4908		
	95%CI for OR	(0.65, 1.09)	(0.73, 4.80)	(0.45, 2.65)	**2.00, 7.10)**	(0.63, 2.59)		
Heart Disease	OR	**0.68**			**1.95**			
	p-value	**0.0016**			**0.0322**			
	95%CI for OR	**(0.54, 0.87)**			**(1.06, 3.58)**			
Cancer	OR	0.71			1.79			
	p-value	0.1853			0.2326			
	95%CI for OR	(0.42, 1.18)			(0.69, 4.64)			
Intestinal Ulcers	OR	**0.56**	1.12		**2.24**			
	p-value	**0.0078**	0.8889		**0.0037**			
	95%CI for OR	**(0.37, 0.86)**	(0.23, 5.42)		**(1.30, 3.85)**			
Bowel Disorder	OR	0.73			1.78			
	p-value	0.3189			0.2677			
	95%CI for OR	(0.39, 1.35)			(0.64, 4.94)			
Osteoporosis	OR	0.89		1.05	1.18	0.65		
	p-value	0.2855		0.9063	0.7990	0.4133		
	95%CI for OR	(0.71, 1.10)		(0.50, 2.21)	(0.34, 4.11)	(0.24, 1.81)		
Long-term	OR	**0.83**	**0.61**	**0.47**	1.14	**0.66**	0.65	0.78
	p-value	**0.0035**	**0.0196**	**0.0001**	0.3487	**0.0181**	0.3217	0.5107
	95%CI for OR	**(0.73, 0.94)**	**(0.41, 0.92)**	**(0.32, 0.69)**	(0.87, 1.48)	**(0.46, 0.93)**	(0.28, 1.52)	(0.38, 1.63)
Chronic	OR	**0.74**	0.84	**0.68**	**1.42**	**0.73**	0.56	0.78
	p-value	**0.0000**	0.3993	**0.0099**	**0.0117**	**0.0439**	0.1351	0.4263
	95%CI for OR	**(0.64, 0.85)**	(0.56, 1.26)	**(0.51, 0.91)**	**(1.08, 1.86)**	**(0.53, 0.99)**	(0.26, 1.20)	(0.42, 1.44)

Table 4.10: Odds Ratios from the Logistic Regression of Long-term Health Conditions and Potato Consumption, Race/Ethnicity

Outcome of Interest		FVCDDPOT	Black	KFJC	AbNA	Asian	Latin American	Other
Blood Pressure	OR	1.07	**2.09**	1.18	**1.87**	0.97		0.45
	p-value	0.4995	**0.0553**	0.4731	**0.0106**	0.8873		0.0785
	95%CI for OR	(0.89, 1.28)	**(0.98, 4.45)**	(0.75, 1.85)	**(1.16, 3.03)**	(0.60, 1.57)		(0.19, 1.10)
Diabetes	OR	0.87	1.83	1.08	**3.98**	1.25		
	p-value	0.3055	0.2002	0.8742	**0.0000**	0.5345		
	95%CI for OR	(0.67, 1.13)	(0.72, 4.64)	(0.44, 2.64)	**(2.14, 7.42)**	(0.62, 2.52)		
Heart Disease	OR	1.26			**2.00**			
	p-value	0.1040			**0.0276**			
	95%CI for OR	(0.95, 1.66)			**(1.08, 3.69)**			
Cancer	OR	**1.56**			1.70			
	p-value	**0.0300**			0.2803			
	95%CI for OR	**(1.04, 2.33)**			(0.65, 4.47)			
Intestinal Ulcers	OR	1.41	1.17		**2.25**			
	p-value	0.0747	0.8431		**0.0048**			
	95%CI for OR	(0.97, 2.08)	(0.24, 5.68)		**(1.28, 3.96)**			
Bowel Disorder	OR	0.77			1.90			
	p-value	0.2152			0.2199			
	95%CI for OR	(0.51, 1.17)			(0.68, 5.28)			
Osteoporosis	OR	0.98		1.06	1.20	0.66		
	p-value	0.8784		0.8870	0.7780	0.4135		
	95%CI for OR	(0.76, 1.27)		(0.50, 2.25)	(0.34, 4.17)	(0.24, 1.81)		
Long-term	OR	0.93	**0.61**	**0.47**	1.17	**0.65**	0.64	0.76
	p-value	0.3679	**0.0172**	**0.0002**	0.2355	**0.0129**	0.3118	0.4721
	95%CI for OR	(0.80, 1.08)	**(0.40, 0.92)**	**(0.32, 0.70)**	(0.90, 1.53)	**(0.46, 0.91)**	(0.26, 1.51)	(0.37, 1.59)
Chronic	OR	1.14	0.85	**0.74**	**1.45**	**0.72**	0.54	0.76
	p-value	0.1061	0.4364	**0.0430**	**0.0058**	**0.0399**	0.1325	0.3447
	95%CI for OR	(0.97, 1.34)	(0.57, 1.28)	**(0.55, 0.99)**	**(1.11, 1.89)**	**(0.53, 0.99)**	(0.24, 1.21)	(0.43, 1.35)

Table 4.11: Odds Ratios from the Logistic Regression of Long-term Health Conditions and Carrot Consumption, Race/Ethnicity

Outcome of Interest		FVCDDCAR	Black	KFJC	AbNA	Asian	Latin American	Other
Blood Pressure	OR	0.97	2.06	1.15	**1.91**	0.96		0.45
	p-value	0.7649	0.0602	0.5312	**0.0084**	0.8561		0.0761
	95%CI for OR	(0.77, 1.21)	(0.97, 4.38)	(0.74, 1.80)	**(1.18, 3.09)**	(0.59, 1.55)		(0.19, 1.09)
Diabetes	OR	0.81	1.89	1.11	**3.75**	1.27		
	p-value	0.1486	0.1836	0.8248	**0.0000**	0.4999		
	95%CI for OR	(0.61, 1.08)	(0.74, 4.82)	(0.45, 2.72)	**(1.99, 7.07)**	(0.63, 2.57)		
Heart Disease	OR	0.89			**2.09**			
	p-value	0.4051			**0.0162**			
	95%CI for OR	(0.67, 1.17)			**(1.15, 3.82)**			
Cancer	OR	0.87			1.84			
	p-value	0.4877			0.2035			
	95%CI for OR	(0.59, 1.28)			(0.72, 4.74)			
Intestinal Ulcers	OR	**0.42**	1.15		**2.27**			
	p-value	**0.0015**	0.8604		**0.0035**			
	95%CI for OR	**(0.25, 0.72)**	(0.24, 5.58)		**(1.31, 3.92)**			
Bowel Disorder	OR	**0.60**			1.78			
	p-value	**0.0152**			0.2730			
	95%CI for OR	**(0.40, 0.91)**			(0.64, 4.95)			
Osteoporosis	OR	0.89		1.05	1.18	0.65		
	p-value	0.4153		0.8916	0.7975	0.4100		
	95%CI for OR	(0.66, 1.19)		(0.50, 2.24)	(0.34, 4.10)	(0.24, 1.81)		
Long-term	OR	**0.84**	**0.62**	**0.47**	1.16	**0.65**	0.65	0.77
	p-value	**0.0332**	**0.0208**	**0.0002**	0.2831	**0.0141**	0.3179	0.4900
	95%CI for OR	**(0.72, 0.99)**	**(0.41, 0.93)**	**(0.32, 0.70)**	(0.89, 1.51)	**(0.46, 0.92)**	(0.28, 1.52)	(0.37, 1.61)
Chronic	OR	0.90	0.84	**0.71**	**1.46**	**0.71**	0.55	0.76
	p-value	0.1692	0.3910	**0.0201**	**0.0055**	**0.0308**	0.1337	0.3634
	95%CI for OR	(0.77, 1.05)	(0.56, 1.25)	**(0.53, 0.95)**	**(1.12, 1.91)**	**(0.52, 0.97)**	(0.25, 1.20)	(0.42, 1.37)

Table 4.12: Odds Ratios from the Logistic Regression of Long-term Health Conditions and Other Vegetables Consumption, Race/Ethnicity

Outcome of Interest		FVCDDVEG	Black	KFJC	AbNA	Asian	Latin American	Other
Blood Pressure	OR	0.99	2.05	1.16	**1.87**	0.96		0.45
	p-value	0.8755	0.0607	0.5213	**0.0116**	0.8558		0.0763
	95%CI for OR	(0.92, 1.08)	(0.97, 4.35)	(0.74, 1.82)	**(1.15, 3.05)**	(0.59, 1.55)		(0.19, 1.09)
Diabetes	OR	**0.86**	1.82	1.18	**3.74**	1.28		
	p-value	**0.0370**	0.2155	0.7140	**0.0001**	0.4943		
	95%CI for OR	**(0.75, 0.99)**	(0.71, 4.67)	(0.49, 2.83)	**(1.98, 7.08)**	(0.63, 2.58)		
Heart Disease	OR	0.89			**1.96**			
	p-value	0.0661			**0.0369**			
	95%CI for OR	(0.78, 1.01)			**(1.04, 3.68)**			
Cancer	OR	0.93			1.87			
	p-value	0.2946			0.1981			
	95%CI for OR	(0.81, 1.07)			(0.72, 4.84)			
Intestinal Ulcers	OR	**0.81**	1.08		**2.28**			
	p-value	**0.0278**	0.9267		**0.0035**			
	95%CI for OR	**(0.67, 0.98)**	(0.22, 5.18)		**(1.31, 3.97)**			
Bowel Disorder	OR	0.96			1.83			
	p-value	0.6340			0.2438			
	95%CI for OR	(0.83, 1.12)			(0.66, 5.05)			
Osteoporosis	OR	1.02		1.06	1.21	0.66		
	p-value	0.6852		0.8874	0.7666	0.4190		
	95%CI for OR	(0.91, 1.15)		(0.50, 2.24)	(0.35, 4.21)	(0.24, 1.82)		
Long-term	OR	**0.94**	**0.60**	**0.49**	1.14	**0.65**	0.67	0.77
	p-value	**0.0378**	**0.0161**	**0.0004**	0.3222	**0.0153**	0.3540	0.4694
	95%CI for OR	**(0.88, 1.00)**	**(0.40, 0.91)**	**(0.33, 0.73)**	(0.88, 1.49)	**(0.46, 0.92)**	(0.28, 1.57)	(0.37, 1.58)
Chronic	OR	**0.89**	0.81	**0.74**	**1.43**	**0.72**	0.56	0.75
	p-value	**0.0002**	0.3134	**0.0408**	**0.0082**	**0.0323**	0.1543	0.3402
	95%CI for OR	**(0.84, 0.95)**	(0.55, 1.21)	**(0.55, 0.99)**	**(1.10, 1.86)**	**(0.52, 0.97)**	(0.25, 1.24)	(0.42, 1.35)

Table 4.13: Goodness of Fit Tests (p-value) for Logistic Models with Total Fruit and Vegetables

	P	D	HL	Osius	McC
Model: logit(blood pressure) = -5.70 + 0.07age + 0.08sex - 0.10 _5ormoreserv 0.73Black + 0.14KFJC + 0.62AbNA - 0.05Asian -0.79Other	1658.19 (0.01)	1643.35 (0.02)	295.65 (0.01)	1.29 (0.10)	1.38 (0.08)
Model: logit(diabetes) = -6.09 + 0.06age + 0.25sex – 0.28 _5ormoreserv 0.64Black + 0.09 KFJC + 1.28 AbNA + 0.23 Asian	1421.83 (0.05)	1158.53 (0.99)	167.79 (0.00)	0.74 (0.23)	0.81 (0.21)
Model: logit(heart disease) = -7.33 + 0.07age + 0.54sex – 0.16 _5ormoreserv + 0.69AbNA	730.74 (0.01)	626.63 (0.55)	66.06 (0.01)	1.24 (0.11)	1.31 (0.10)
Model: logit(cancer) = -6.96 + 0.05 age + 0.22 sex – 0.05 _5ormoreserv + 0.67AbNA	592.66 (0.87)	510.88 (0.99)	15.69 (0.05)	-0.49 (0.69)	-0.43 (0.67)
Model: logit(intestinal ulcers) = -4.98 + 0.03age – 0.03sex – 0.18 _5ormoreserv + 0.12Black + 0.85AbNA	844.18 (0.26)	703.15 (0.99)	56.83 (0.01)	0.28 (0.39)	0.36 (0.36)
Model: logit(bowel disorder) = -4.60 + 0.03age – 1.01 sex + 0.06 _5ormoreserv + 0.63AbNA	629.08 (0.53)	542.94 (0.99)	39.01 (0.01)	-0.10 (0.54)	-0.04 (0.52)
Model: logit(osteoporosis) = -4.67 + 0.05age – 2.15sex + 0.01 _5ormoreserv + 0.06KFJC + 0.19AbNA – 0.42Asian	514.83 (0.01)	423.18 (0.60)	24.46 (0.02)	1.50 (0.07)	1.63 (0.05)
Model: logit(longterm) = -1.79 + 0.02age – 0.25sex – 0.12 _5ormoreserv -0.49Black – 0.74KFJC + 0.13AbNA – 0.44Asian -0.38Latin_american – 0.26Other	1847.67 (0.00)	1967.10 (0.00)	117.05 (0.00)	5.23 (0.00)	5.40 (0.00)
Model: logit(chronic) = -2.27 + 0.05age -0.26sex – 0.15 _5ormoreserv -0.18Black -0.34KFJC - 0.35Asian – 0.53Latin_american - 0.27Other	1953.67 (0.00)	2096.82 (0.00)	175.67 (0.00)	6.85 (0.00)	7.03 (0.00)

P: Pearson Statistic   *D: Deviance Statistic*   *HL: Hosmer Lameshow Goodness of Fit Test*
O: Osius-Rojek Test   *McC: McCullagh Test*

Chapter Five:

Epilogue and Conclusions

5.1 Introduction

This chapter reviews the core findings of this research in *Section 5.2*. Concluding remarks, limitations of this study, further research recommendations make up *Section 5.3*.

5.2 Discussion of Major Findings

Major Findings:

(i) The majority of people in all ethnic groups consume less than the Health Canada recommended fruit and vegetable benchmark '5 or more times/servings of fruit and vegetables per day'. Substantially low percentages of Southeast Asian, Aboriginal people of North America, West Asian, Korean, Japanese and Chinese ethnic groups compared to other racial/cultural origins take 5 or more times/servings of fruit and vegetables per day.

(ii) The summary chart in Tabel 5.1 delineates the significant associations between fruit & vegetables and long-term diseases derived from logistic regression runs using 500 bootstrap replicates. The symbol ↑ and ↓ indicates increasing and decreasing propensity respectively. As for instance, in the first case, having 5 or more times/servings of fruit and vegetables per day is found be to be associated with lower diabetes, other long-term health conditions and chronic health conditions.

Table 5.1: Summary Chart of Association between Fruit & Vegetables and Long-term Diseases

FVI	Direction of Association with Long-term Diseases
Total Fruit and Vegetables (5 or more times/servings per day) ↑	↓ Diabetes ↓ Other Long-term Health Conditions ↓ Chronic Health Conditions
Fruit Juice ↑	↓ Diabetes
Fruit ↑	↓ Heart Disease ↓ Other Long-term Health Conditions ↓ Chronic Health Conditions
Green Salad ↑	↓ Heart Disease ↓ Intestinal Ulcers ↓ Other Long-term Health Conditions ↓ Chronic Health Conditions
Potato ↑	↑Cancer
Carrot ↑	↓ Intestinal Ulcers ↓ Bowel Disorders ↓ Other Long-term Health Conditions
Other Vegetables ↑	↓ Diabetes ↓ Intestinal Ulcers ↓ Bowel Disorders ↓ Other Long-term Health Conditions

(iii) The summary chart in Table 5.2 indicates the direction of propensity of having long-term diseases for each ethnic group compared to the reference ethnic group, *White*. As for instance, in the first case, *Black* racial groups are found to have less propensity of having high blood pressure and other long-term health conditions compared to *White* ethnic groups.

Table 5.2: Summary Chart of Association between Ethnicity and Long-term Diseases

Ethnic Groups	Direction of Propensity of having long-term diseases (compared to *White* ethnic group)
Black	↓ High Blood Pressure ↓ Other Long-term Health Conditions
KFJC (Korean, Filipino, Japanese, Chinese)	↓ Other Long-term Health Conditions ↓ Chronic Conditions
Aboriginal People of North America	↑ High Blood Pressure ↑ Diabetes ↑ Heart Disease ↑ Intestinal Ulcers ↑ Chronic Health Conditions
Asian	↓ Other Long-term Health Conditions ↓ Chronic Conditions
Latin American	No Significant Association
Other	No Significant Association

(iv) Human species are often classified to discover 'novel' risk factors for diseases. This classification implies that people within each racial group share common cultural characteristics such as language, diet, and social networks, and who may have an increased prevalence of specific genotypes (60). Therefore, studies of disease etiology should incorporate ethnic differences. Critics of research aimed at determining ethnic differences in health believe that, while abundant, much of the ethnicity research has not led to new understandings about diseases, and label this type of research as 'black box epidemiology' (61). Black box epidemiology refers to epidemiology where the exact causal mechanism behind the association remains hidden (black), but the inference is that the causal association may be found within the association (box). The black box approach of ethnicity based research is not deemed useful as it does not discover etiology but usually generates potential clues to the pathogenesis of many diseases. Several logistic regressions have been run in this analysis to estimate the propensity of having long-term diseases for all racial groups with respect to White group (reference group). Another facet of this study is to examine the association of 5 or more times/servings of fruit and vegetables per day on the long-term diseases. As per 'black box epidemiology', 'FV consumption' and 'ethnicity' bridge together and act against the backdrop of the pathogenesis of long-term diseases. As assessed by CCHS 2.2, Aboriginal people of North America tend to exhibit higher susceptibility to long-term diseases as opposed to White ethnic group which might be due to least intake of fruit and vegetables consumption.

(v) CCHS 2.2 poses the problem of sparse data while estimating goodness of fit test in logistic model. Majority of the logistic models are appraised as well-fit by Deviance statistic, Osius-Rojek and McCullagh goodness of fit test. The problem of sparsity becomes intensified in case of several covariates and all goodness of fit tests fail.

5.3 Finale

This study emanates from the premise that lower intake of fruit and vegetables is an underlying or exacerbating cause of long-term disease prevalence in Canadian ethnic groups. However, this research has some limitations that merit further discussion. It might be argued that ethnic comparisons may be confounded by social factors. Differences in social factors produce inequalities in exposure to, susceptibility to, and resistance from pathogenic processes across the life span (62,63). Therefore, studies of disease etiology should incorporate both social and biologic factors, and ethnic differences must be viewed as possibly being due to differences in biology, social factors, or both. Failure to address the social dimension of the ethnicity construct results in misinterpretation of the research results. Here, ethnic comparisons have been adjusted for age and sex but other socioeconomic disparities viz. income, employment status might be incorporated which would lead to more subtle ethnic differences. In the same fashion, disease burden in all ethnic groups may be explicated through the fruit and vegetables covariate adjusted for potential confounders viz. smoking, physical activity, and consumption of alcohol which are conducive to long-term diseases. But nonetheless, the positive spin-off of this research obviously outweighs the negative spin-off as it clearly demonstrates emphatic clues to the pathogenesis of the long-term diseases in all ethnic groups. As to the final recommendations, it can be said unequivocally that further research will receive impetus from such ethnicity based research. The nutrients intake estimation in different ethnic groups as well as their within and between group differences might intrigue exploring multi-ethnic population's nutritional status. The important nutrients absorption through fruit and vegetables may be explored by Statistics Canada recommended software

'Software for Intake Distribution Estimation' (SIDE) in estimating nutritional status of all ethnic groups in Canada.

References

1. Sheth, T et al. Classifying ethnicity utilizing the Canadian Mortality database. *Ethnicity and Health*. 1997; 2(4): 287-295.

2. *American Cancer Society. Cancer Facts & Figures 2003.* Atlanta, Ga: American Cancer Society; 2003.

3. Crawford ED. Epidemiology of prostate cancer. *Urology*. 2003; 62: 3-12.

4. Centers for Disease Control and Prevention. *The burden of chronic diseases and their risk factors: National and state perspectives, 2004.* Available at http://www.cdc.gov/nccdphp/burdenbook2004.

5. Hung H., Joshipura K. J., et al. Fruit and vegetable intake and risk of major chronic disease. *Journal of the National Cancer Institute*. 2004; 96(21): 1577-1584.

6. Genkinger J. M., Platz E. A., Hoffman S. C., Comstok G. W. and Helzlsouer K. J. Fruit, vegetable and antioxidant intake and all-cause, cancer, and cardiovascular disease mortality in a community-dwelling population in Washington county, Maryland. *American Journal of Epidemiology*. 2004; 160(12): 1223-1233.

7. Adebawo O. et al. Fruit and vegetables moderate lipid cardiovascular risk factor in hypertensive patients. *Lipids in Health and Disease*. 2006; 5: 14-18.

8. Canadian Food Guide. Health Canada. 2007; Available from: *http://www.hc-sc.gc.ca/fn-an/food-guide-aliment/index_e.html*

9. *Statistics Canada. Canadian Community Health Survey: A First Look.* Available from: www.statcan.ca/Daily/English/020508/d020508a.htm.

10. Cooper, R. A note on the biologic concept of race and its application in epidemiologic research. *American Heart Journal.* 1984; 108: 715-722.

11. Claudio E. P. Fruit and vegetable consumption. *Statistics Canada, Catalogue 82-003, Health Reports*. 2002; 13: 23-31.

12. Ling A. M. C. and Horwath C. Perceived benefits and barriers of increased fruit and vegetable consumption: Validation of a decisional balance scale. *Journal of Nutrition Education*. 2001; 33(5): 257-265.

13. Blanchette L. and Brug J. Determinants of fruit and vegetable consumption among 6-12-year-old children and effective interventions to increase consumption. *Journal of Human Nutrition and Dietetics*. 2005; 18, 431-443.

14. Mette R., Rikke K., Knut-Inge K., Leslie L., Johannes B., Elling B. and Pernille D. Determinants of fruit and vegetable consumption among children and adolescents: A review of the literature. Part I: quantitative studies. *International Journal of Behavioral Nutrition and Physical Activity*. 2006; 3: 22

15. Lindstrom M., Hanson B. S., Wirfait E. and Ostergren P. O. Socioeconomic differences in the consumption of vegetables, fruit and fruit juices. *European Journal of Public Health*. 2001: 11: 51-59.

16. Moser R. P., Green V., Weber D. and Doyle C. Psychosocial correlates of fruit and vegetable consumption among African American Men. *Journal of Nutrition Education and Behaviour*. 2005; 37: 306-314.

17.Anand, S. Ethnicity and the determinants of cardiovascular disease among South Asians, Chinese, and European Canadians. *Ph.D Dissertation (HRM), McMaster University. 2002.*

18. Peeters P. H. M. et al. *Consumption of Vegetables and Fruits and Risk of Breast Cancer. 2005*; 293: 183-193.

19. Olsen A., Tjonneland A., Thomsen B. L., Loft S., Stripp C., Overvad K., Moller S. and Olsen J. H. Fruits and vegetable intake differentially affects estrogen receptor negative and positive breast cancer incidence rates. *The Journal of Nutrition*. 2003; 133: 2342-2347.

20. Castelli W. P. Cholesterol and lipids in the risk of coronary artery disease – the Framingham Heart Study. *Canadian Journal of Cardiology*. 1988; 4(suppl): 5A-10A.

21. Castelli W. P., Anderson K. M., Wilson P. W. F. and Levy D. Lipids and risk of coronary heart disease. The Framingham Study. *Annals of. Epidemiology*. 1992; 2: 23-28.

22. Djousse L., Arnett D. K., Coon H., Province M. A., Moore L. L. and Ellison R. C. Fruit and vegetable consumption and LDL cholesterol: the National Heart, Lung, and Blood Institute Family Heart Study. *American Journal of Clinical Nutrition*. 2004; 79: 213-217.

23. Rissanen T. H., Voutilainen S. and Virtanen J. K., et al. Low intake of fruits, berries and vegetables is associated with excess mortality in men: The Kuopio Ischaemic Heart Disease Risk Factor (KIHD) Study. *Journal of Nutrition*. 2003; 133: 199-204.

24. Johnsen S. P., Overvad K. and Stripp C. et al. Intake of fruit and vegetables and the risk of ischemic stroke in a cohort of Danish men and women. *The American Journal of Clinical Nutrition*. 2003; 78: 57-64.

25. Ziegler R. G. Vegetables, fruits, and carotenoids and the risk of cancer. *The American Journal of Clinical Nutrition*. 1991; 53: 251S – 9S.

26. Riboli E. and Norat T. Epidemiologic evidence of the protective effect of fruit and vegetables on cancer risk. *The American Journal of Clinical Nutrition*. 2003; 78(suppl): 559S-569S.

27. Christen W. G., Liu S., Schaumberg D. A. and Buring J. E. Fruit and vegetable intake and the risk of cataract in women. *The American Journal of Clinical Nutrition*. 2005; 81: 1417-1422.

28. Velde S. J., Wind M., Lenthe F. J., Klepp K. and Brug J. Differences in fruit and vegetable intake and determinants of intakes between children of Dutch origin and non-Western ethnic minority children in the Netherlands – a cross sectional study. *International Journal of Behavioral Nutrition and Physical Activity*. 2006; 3: 31-43.

29. Park S., Murphy S., Wilkens L. R., Yamamoto J. F., Sharma S., Hankin J. H., Handerson B. E. and Kolonel L. N. Dietary patterns using the food guide pyramid groups are associated with sociodemographic and lifestyle factors: The multiethnic cohort study. *The Journal of Nutrition*. 2005; 135: 843-849.

30. Reynolds K. D., Baranowski T., Bishop D. B., Farris R. P., Binkley D., Nicklas T. A. and Elmer P. J. Patterns in child and adolescent consumption of fruit and vegetables:

Effects of gender and ethnicity across four sites. *Journal of American College of Nutrition*. 1999; 18: 248-254.

31. Lindquist C. H., Gower B. A. and Goran M. I. Role of dietary factors in ethnic differences in early risk of cardiovascular disease and type 2 diabetes. *The American Journal of Clinical Nutrition*. 2000; 71: 725-732.

32. Haire-Joshu D., Kreuter M. K., Holt C. and Steger-May K. Estimates of fruit and vegetable intake in childhood and adult dietary behaviors of African American women. *Journal of Nutrition Education and Behaviour*. 2004; 36(6):309-314.

33. *National Institutes of Health. Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults.* Bethesda, Md: National Institutes of Health; 1998. Report No.: 98-4083.

34. *National Institutes of Health. National diabetes information clearinghouse. National Diabetes Statistics.* Bethesda, Md: National Institutes of Health; 2002. Report No.: 02-3892.

35. Moser R. P., Green V., Weber D. and Doyle C. Psychosocial correlates of fruit and vegetable consumption among African American men. *Journal of Nutrition Education and Behaviour*. 2005; 37(6): 306-314.

36. Devine C. M., Wolfe. W. S., Frongillo E. A., and Bisogni C. A. Life-course events and experiences: Association with fruit and vegetable consumption in three ethnic groups. *Journal of the American Dietetic Association*. 1999; 99(3): 309-312.

37. Li R., Serdula M., Bland S., Mokdad A., Bowman B. and Nelson D. Trends in fruit and vegetable consumption among adults in 16 US states: Behavioral risk factor

surveillance system, 1990-1996. *American Journal of Public Health.* 2000; 90(5): 777-781.

38. Cotugna N. and Fleming E. Comparison of standard and portion-assisted responses to a brief fruit and vegetable food frequency questionnaire in a Hispanic WIC population. *Journal of the American Dietetic Association.* 1998; 98(12): 1462-1464.

39. Maskarinec G., Novotny R. and Tasaki K. Dietary patterns are associated with body mass index in multiethnic women. *The Journal of Nutrition.* 2000; 130: 3068-3072.

40. Cullen K. W., Baranowski T. and Owens E. et al. Ethnic differences in social correlates of diet. *Health Education Research.* 2002; 17(1): 7-18.

41. Bingham S. A. et al. Is 'five-a-day' an effective way of increasing fruit and vegetable intakes? *Public Health Nutrition. 2003*; 7(2): 257-261.

42. Jonneland A., Haraldsdottir J. and Overvad K. et al. Influences of individually estimated portion size data on the validitiy of a semiquantitative food frequency questionnaire. *International Journal of Epidemiology.* 1992; 21: 770-777.

43. Samet J., Humble C. and Skipper B. Alternatives in the collection and analysis of food frequency interview data. *American Journal of Epidemiology.* 1984; 120: 572-581.

44. Block G. A review of validations of dietary assessment methods. *American Journal of Epidemiology.* 1982; 115: 492-498.

45. Clapp J., McPherson R., Reed D. and Hsi B. Comparison of a food frequency questionnaire using reported vs standard portion sizes for classifying individuals according to nutrient intake. *Journal of American Dietetic Association.* 1991; 91: 316-320.

46. Campbell M. K., Polhamus B and. et al. Assessing fruit and vegetable consumption in a 5 a day study targeting rural blacks: The issue of portion size. *Journal of the American Dietetic Association*. 1996; 96: 1040-1042.

47. Pomerleau J., Lock K., McKee M. and Altmann D. R. The challenge of measuring global fruit and vegetable intake. *The Journal of Nutrition*. 2004; 134: 1175-1180.

48. Bankier, M. Power Allocations: Determining sample sizes for subnational areas. *The American Statistician*. 1988; 42: 174-177.

49. *Canadian Community Health Survey Cycle 2.2, Nutrition (2004): A Guide to Accessing and Interpreting the Data*. Health Canada, p.36

50. Lohr, S. L. *Sampling: Design and Analysis. Duxbury Press*. 1999.

51. Rao, J. N. K. and Wu, C. F. J. Resampling inference with complex survey data. *Journal of the American Statistical Association*. 1988; 83: 231-241.

52. Rao, J. N. K., Wu, C. F. J. and Yue, K. Some recent work on resampling methods for complex surveys. *Survey Methodology*. 1992; 18: 209-217.

53. Koch, G. G. and Edwards, S. Logistic regression. *Encyclopaedia of Statistical Sciences*. 1985; Vol V, John Wiley and Sons, New York.

54. Kleinbaum, D. G. and Klein, M. *Logistic Regression: A Self-Learning Text*. Springer. 2002.

55. Kuss, Oliver. Global Goodness-of-fit tests in logistic regression with sparse data. *Statistics in Medicine*. 2002; 21: 3789-3801.

56. Hosmer, D. W. and Lemeshow, S. A goodness of fit test for the multiple logistic regression model. *Communications in Statistics*. 1980; A10: 1043-1069.

57. Osius, G. and Rojek, D. Normal Goodness-of-Tests for multinomial models with large degrees of freedom. *Journal of the American Statistical Association*. 1992; 87: 1145-1152.

58. McCullagh, P. On the asymptotic distribution of Pearson's statistic in linear exponential family models. *International Statistical Review*. 1985; 53: 61-67.

59. Farrington, C. P. On assessing goodness of fit of generalized linear models to sparse data. *Journal of the Royal Statistical Society*. 1996; B, 58: 349-360.

60. Aspinall, P. J. The conceptual basis of ethnic group terminology and classifications. *Social Science & Medicine*. 1997; 45(5): 689-698.

61. Senior, P. A. and Bhopal, R. Ethnicity as a variable in epidemiological research. *British Medical Journal*. 1994; 309: 327-330.

62. Krieger, N. Analyzing socioeconomic and racial/ethnic patterns in health and health care. *American Journal of Public Health*. 1993; 83(8): 1086-1087.

63. Ben-Shlomo, Y. and Smith, G.D. Commentary: socioeconomic position should be measured accurately. *British Medical Journal*. 1999; 318(7187): 844-845.