THE ASSOCIATION BETWEEN BEVERAGE INTAKE AND OVERWEIGHT AND OBESITY AMONG CANADIAN ADULTS

A Thesis Submitted to the College of

Graduate Studies and Research

in Partial Fulfillment of the Requirements

for the Degree of Master of Science

in the Department of Community Health and Epidemiology

College of Medicine

University of Saskatchewan

Saskatoon, Saskatchewan

By

Nooshin Nikpartow

PERMISSION TO USE

In presenting this thesis/dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis/dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Head of the Department of Community Health and Epidemiology
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5E5
Canada

OR

Dean
College of Graduate Studies and Research
University of Saskatchewan
107 Administration Place
Saskatoon, Saskatchewan S7N 5A2
Canada

ABSTRACT

Overweight and obesity in Canada has significantly increased during the last three decades, paralleled by increased intake of fat and sugar particularly sugary beverages leading to higher level of energy intake, as well as reduction in physical activity. Canadian Community Health Survey, Cycle 2.2, 2004 (CCHS 2.2), provides the opportunity to evaluate beverage intakes of Canadians in relation to overweight and obesity using Body Mass Index (BMI).

To examine the association between sugar-sweetened beverages and BMI in Canadian adults, we used data from CCHS 2.2 (n=14,304, aged >18 year and <65 year) in which dietary intake was assessed using 24-h recall. In various steps, data on beverage consumption were identified, coded and classified. Using descriptive statistics, we determined total gram intake and the contribution of each beverage to total energy intake among age/sex groups. To determine the most suitable patterns of beverage consumptions among Canadian adults, a cluster analysis K-means method was applied. Males and females were classified into distinct clusters based on the dominant pattern of beverage intakes. Finally, step-wise logistic regression models were used to determine associations between sugar-sweetened beverages and BMI, controlling for age, marital status, income, education, physical activity, total energy intake, immigration status, smoking habits and ethnicity. To account for complex survey design, all data were weighted and bootstrapped.

BMI in women with predominant “fruit drink” pattern (791.1±32.9 g) was significantly higher than those with no dominant pattern in beverage consumption (28.3±1 vs. 26.8±0.3 respectively, P<0.001). In women, high intake of fruit drinks was a significant predictor of overweight (OR=1.84, 95% C.I:1.06-3.20), obesity (OR=2.55, 95% C.I:1.46-4.47) and overweight/obesity (OR=2.05, 95% C.I:1.29-3.25). In men, mean BMI was not different among beverage consumption clusters and none of the beverages was a predictor for overweight and obesity. For the first time, in a nationally representative data, we report association of sugar-sweetened beverages and overweight and obesity in Canadian women.

ACKNOWLEDGEMENTS

I am grateful for the financial support received from the Canadian Institute of Health Research (CIHR) while studying in the Department of Community Health & Epidemiology and working on this thesis project.

There are many people to whom I owe many thanks for their support and guidance throughout my academic career and this thesis journey.

To my supervisors, Dr. Hyun Lim, for her methodological advice and constantly questioning and challenging my thoughts; and Dr. Hassan Vatanparast, for his unending encouragement and support through the rough times and his invaluable insights in the development of this project. Words are not able to express my gratitude.

I am also grateful to my advisory committee members Dr. Bonnie Janzen and Dr. Susan Whiting for their long-standing involvement and valuable comments.

I am also thankful to my co-researcher, Adrienne Danyliw, for her cooperation and her thought provoking and engaging discussion.

This project would not have been achievable without the support and partnership of Saskatchewan Research Data Centre (SKY RDC) Statistics Canada and statistical analyst Mr. Jesse McCrosky, as well as statistical assistant Mr. Lloyd Balbuena. I am very grateful for the warm and welcoming environment provided for me at SKY RDC, and for the precious statistical hints given to me.

Finally, I wish to thank my husband, for his patience and emotional support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACE	Angiotensin-converting Enzyme
ANOVA	Analysis of Variance
AMPM	Automated Multiple-Pass Method
BNS	Bureau of Nutritional Sciences
BMI	Body Mass Index
BMR	Basal Metabolic Rate
BRR	Balanced Repeated Replication
BW	Body Weight
CA	Cluster Analysis
CAI	Computer-assisted Interviewing
CCC	Cubic Clustering Criterion
CCHS 2.2	Canadian Community Health Survey Cycle 2.2
CDC	Centers for Disease Control and Prevention
CI	Confidence Interval
CNF	Canadian Nutrient File
CSFII	Continuing Survey of Food Intakes by Individuals
CT	Computerized Tomography
CV	Coefficients of Variations
DRI	Dietary Reference Intakes
F	Females
FDC	Food Description
FFQ	Food Frequency Questionnaires
FID	Food and Ingredients Details
FRL	Food Recipe Level
HF	High-fat
HFCS	High-fructose Corn Syrup
HFSSM	Household Food Security Survey Model
HS	Health Survey
IOM	Institute of Medicine
LF	Low-fat
M	Males
NHANES	National Health and Nutrition Examination Survey
NSS	Nutrition Survey System
OHS	Ontario Health Survey
OR	Odds Ratio
PCA	Principle Component Analysis
PFS	Pseudo-F statistic
R24	Recall 24-hour
RRR	Reduced Rank Regression
SD	Standard Deviation
SEM	Standard Error of the Mean
SSB	Sugar-sweetened Beverages
US	United States
WHO	World Health Organization

Chapter 1

INTRODUCTION

1.1 Background Information and Rationale

The prevalence of overweight and obesity in North America has significantly increased during the last three decades.[1] Obesity and overweight are often described as excessive or abnormal fat accrual in adipose tissue, contributing to the development and progression of cardiovascular diseases, Type II diabetes, osteoarthritis, gallbladder disease, some cancers, and premature death.[2,3]

The positive trend of overweight and obesity has been paralleled by the global change in dietary patterns leading to higher levels of energy intake, as well as by a notable reduction in levels of physical activity.[4] The increased caloric intake is mostly from higher dietary fat and sugar. In the United States, analysis of National Health and Nutrition Examination Survey (NHANES) data reveals that 15-25% of average daily total energy intake of American adults derives from beverages.[5] In addition, NHANES data also show an increase in the consumption of sugar-sweetened beverages (SSB) over the past 20 years, thus becoming the largest source of beverage calories for adults today.[6]

Concomitant increases in the consumption of sugar-sweetened beverages (SSB) and obesity rates in adult Americans suggest a possible causative association.[5,7] Although the relationship between beverage consumption and weight management in children has been addressed frequently in recent literature,[8] studies focusing on the adult population are lacking. The potential negative health impacts of SSB intake have directed some to advocate a tax for sugar-sweetened beverages,[9] however more solid evidence linking overweight and obesity to higher intake of sweetened beverages is needed to achieve it.[10]

In Canada, the only available nutrition data after Nutrition Canada (1970-72) is the Canadian Community Health Survey Cycle 2.2 (CCHS 2.2), year 2004, which provides information on food intake (including beverage) of Canadians, as well as measured weight and height at both national and provincial levels. In 2008, Didier Garriguet[11] published an initial descriptive analysis of beverage intake of Canadian adults using data from the CCHS 2.2. However, to our knowledge, no studies have explored patterns of beverage intake of Canadian adults or have investigated the association of sugar-sweetened beverages and Body Mass Index (BMI) on a national scale.

1.2 Objectives of the Study

The primary objective of this study is to characterize daily consumption of beverages of Canadian adults according to age groups (19-30 year, 31-50 year, 51-65 year) and sex. In this study, the contribution of sweetened-beverages, milk, alcoholic beverages, and juice to total energy intake of Canadian adults in various age-sex groups will be assessed.

The second objective is to explore the beverage intake pattern of Canadian adults by evaluating and comparing their socio-economic status, demographics, and other characteristics of participants from different beverage clusters. Moreover, the association of sugar-sweetened beverages with BMI will be assessed, taking into account household income, education, total energy intake, physical activity, smoking, ethnicity, marital status, immigration status, season and residential area (urban/rural). Finally, the beverage intake patterns of Canadian adults will be compared to those of American adults from the most recent published data.[5-7]

1.3 Statement of Hypotheses

To our knowledge, there is no research that has examined patterns of beverage intake of Canadian adults in relation to BMI. We test the hypothesis that Canadian adults, whose beverage consumption is predominantly from sweetened beverages, will have higher body mass index than those who do not exhibit this type of sweetened beverage consumption, and they are more likely to be overweight (BMI 25-29.9) or obese (BMI $\geq$ 30) compared to those with no predominant beverage pattern. We also hypothesize that the relationship between BMI and intake of sweetened beverages in Canadian adults is influenced by other factors such as age, sex, ethnicity, household income, education, and household food security, among others.

Chapter 2

REVIEW OF THE LITERATURE

This Chapter provides the basis for the present study by identifying research done to date in the area of beverage intake and its association with overweight and obesity. Background information is initially presented, describing diet-disease relationship and dietary assessment methods. This is followed by definitions and measurements of obesity and overweight. Different groups of beverages and the association of beverage intake and overweight and obesity among the adult population are also reviewed within different study designs.

2.1 Diet-Disease Relationship

Nutritional studies investigate the role of dietary factors in the causation or prevention of illness and promotion of health.[12-15] However, such studies are challenging due to the complex interaction of food components, as well as the influence of multiple factors on dietary intake and metabolism, such as genetics, behavioural characteristics, socioeconomic status, environmental factors, and physical activity.[12,14] Although cross-sectional and even prospective population studies in nutritional epidemiology do not establish causality of diet-disease relationship, they can offer strong inferred evidence of association.[11] Findings from population studies provide valuable basis for reducing potential random and systematic errors, thus enabling quality improvement in clinical practice and in food processing and technology.[12,15,16]

Unlike traditional nutritional assessment methods that focused on the level associated with a nutrient deficiency, multiple levels of nutrient status are now defined by application of new functional tests combined with regular methods of nutritional assessment (i.e. dietary, laboratory, anthropometric, and clinical).[16] In addition, for a correct interpretation of the assessment results, other variables of demographic and socioeconomic status that influence the nutritional status of participants should also be considered.[16]

2.2 Dietary Assessment Methods

Dietary assessment methods must be reasonably accurate and fairly inexpensive. The level of the surveys (national or individual) defines which methods of dietary assessment need to be used, each of them having advantages and disadvantages.[16] Food consumption measurements at the national level are most commonly based on food balance sheets, and occasionally on total diet studies through market basket, duplicate portion studies or individual food items.[16] At the individual level, one of the most common methods for assessing usual intake of a large population is the 24-hour dietary recall, which is easy, quick, and has high compliance, but relies on the respondent's memory.[16] Food Frequency Questionnaires (FFQ) is another method that is rapid and useful in epidemiologic studies; however it is the least accurate.[16]

A recently developed computer-assisted 24-hour recall, the Automated Multiple-Pass Method (AMPM), is an accurate method for the dietary assessment of a large population. The performance and validity of AMPM has been well-established in studies against other gold-standard methods.[17,18] This method employs research-based strategies to improve dietary recall. The five steps of AMPM help the participants to initially self-define their 24-hour day diet, and later remember forgotten foods and beverages during subsequent steps with detailed description, amount, time and occasions, yet with minimal burden of repetition.[19] Moreover, AMPM allows the interviewer to add, delete or change foods anytime during interview.[19]

A single 24-hour recall of a population reflects both within-person and between-person intake variability which is more variable than the usual dietary intake.[19] To estimate nutrient inadequacy in a group, the distribution of the usual intake is required, and this can be obtained by adding a second 24-recall from a representative sub-sample of participants.[19] The addition of a second 24-hour recall will adjust the distribution of the first 24-hour recall by removing the effects of within-person day-to-day variability. [19] Alternatively, while the purpose of assessing dietary intakes for a group of people is to estimate mean intake levels, a single 24-hour recall is almost representative of the usual intake as it is illustrated in Figure 2.1.

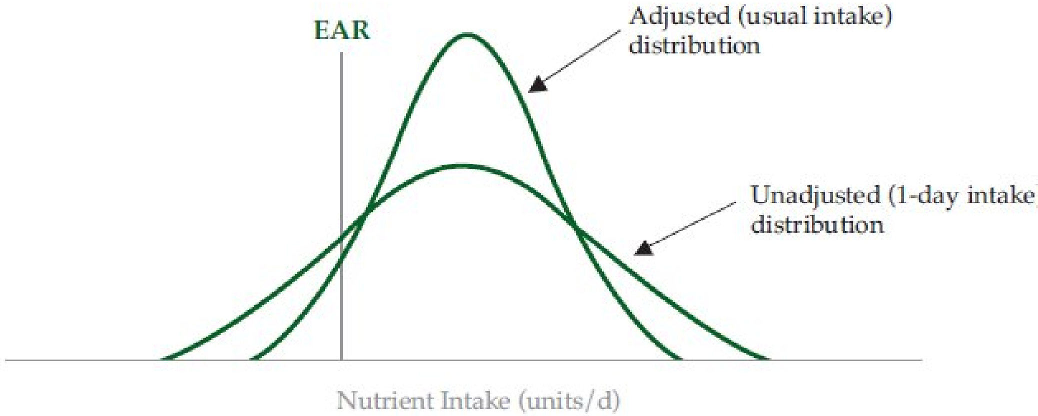


Figure 2.1. Similar means of unadjusted and adjusted intake distributions (Adapted from Stats Canada CCHS 2.2 guidelines) [19]

The potential random and systematic errors that may occur during collection or recording of data vary with the methods used. Different sources of measurement error (i.e. non-response bias, respondent biases including underreporting, interviewer bias, respondent memory lapses, incorrect estimation of portion size, supplement usage, and mistakes in handling of mixed dishes) should be considered before creating the study design and sampling.[16] For example, underreporting that involves both under-recording and under-eating has been documented more in individuals with higher BMI,[20,21] women, and older people.[22]

Random measurement errors impact the reproducibility of the method, and can be reduced by increasing the number of subjects. Systematic errors often lead to an important bias, and can be quantified by calibration studies. To minimize both kinds of measurement errors, different quality-control methods can be used, such as retraining interviewers and coders, standardization and pre-testing the questionnaires, and running a pilot study. It is essential to assess reproducibility and validity of the dietary method, especially for national comparisons.[15]

2.3 Overweight and Obesity

Obesity and overweight are often described as excessive or abnormal fat accrual in adipose tissue, which contributes to health impairment.[2] According to the Centers for Disease Control and Prevention (CDC), these terms are used when the weight for a given height is greater than what is commonly believed as healthy, thus leading to conditions such as high blood pressure, Type II diabetes, cardiovascular diseases, and premature death.[23]

It is well established that excess weight is associated with mortality from cardiovascular diseases[2,3] and cancers,[24] as well as obesity-related morbidity. Obesity-related morbidity can be classified into three categories according to the degree of relative health risks:[25] i)The risk of Type 2 diabetes,[26] insulin resistance,[27] gall bladder disease,[28] dyslipidemia,[29] breathlessness,[30] and sleep apnea,[31] is greatly associated with obesity (relative risk >3); ii) The possibility of coronary heart disease,[32] hypertension,[26] stroke,[33] osteoarthritis,[26] and psychological damage[34] will moderately increase in obesity (relative risk 2-3); iii) There is a slightly increased risk (relative risk 1-2) of different kinds of cancers,[24] hormone abnormalities, impaired fertility,[35] anesthetic risk and fetal defects[26] associated with obesity.

2.3.1 Measurement of Overweight and Obesity

A number of techniques that quantify either total body fat or the fat distribution in the body are available to measure overweight and obesity. The most accurate and advanced method of laboratory measurement is the Computerized Tomography (CT), however it is not economically feasible.[16] Another way to determine obesity is the skin fold caliper which assesses the thickness of skin and subcutaneous fat. This method is adequate in population studies but it is very operator-dependent and it depends on the sex, age, and ethnicity of the subject.[16,36] Other anthropometric methods that are commonly used in clinical practice to estimate risk for obesity-related health problems are based on the location of fat in the body rather than total body fat as with Body Mass Index (BMI).[16] One of the greatest risk factors for cardiovascular diseases is the excess fat in the abdomen which is assumed to be metabolically more active than fat deposition in other parts of the body. Waist circumference and waist-hip ratio are two methods to estimate abdominal fat.[16]

2.3.2 Body Mass Index

Body Mass Index (BMI), determined by measured weight (kg) / measured height^2 (m), is one of the most commonly used indicators of adult population body fatness because of its availability and lower cost.[37] However, BMI application has limitations since it does not account for potential confounders such as age, sex, ethnicity, leg length, and body build.[16,38] Older people (and women) tend to have greater percentage of body fat than younger people (and men) of an equivalent BMI, and muscular professional athletes might have higher BMI with no excess body fat.[16,37] Moreover, the influence of ethnic differences on body composition and fatness has been well-documented, thus showing the importance of assessing such differences in order to avoid misclassification of diverse populations.[36]

Unlike BMI classification for children, there is one consistent BMI classification for adults according to the World Health Organization (WHO), the Centers for Disease Control and Prevention (CDC), and international cut-off points (Table 2.1). Health Canada (2003) warns that the application of this classification system may lead to a misclassification among certain population groups including highly muscular or very lean adults, pregnant women, those over 65 years of age, and certain ethnic groups.[39] However, these factors are less probable to bias the relationship between body fat and BMI in the obese range (BMI ≥ 30).[37]

Table 2.1. Health risk classification according to Body Mass Index (BMI) in adults*

Classification	BMI Category (kg/m^2)	Risk of developing health problems
Underweight	<18.5	Increased
Normal Weight	18.5 - 24.9	Least
Overweight	25.0 - 29.9	Increased
Obese Class I	30.0 - 34.9	High
Obese Class II	35.0 - 39.9	Very high
Obese Class III	≥40.0	Extremely high

*Not for use with pregnant and lactating women.
Note: For persons 65 years and older the 'normal' range may begin slightly above BMI 18.5 and extend into the 'overweight' range.
Adapted from World Health Organization (2000)[39]

2.4 Beverage Categories

A beverage (or a drink) is a liquid particularly prepared for human basic needs, and it is an important part of societal culture. According to the various studies that will be referred to later in this thesis, different types of beverages can be categorized into four main groups: i) sugar-sweetened, ii) nutrient-based, iii) alcoholic, and iv) no additional caloric beverages (including water).

2.4.1 Sugar-sweetened Low Nutrient Beverages

The consumption of sugar-sweetened beverages (SSB) such as fruit drinks (<100% fruit juice from concentrate or fruit flavoured beverages with added sugar) and regular soft drinks (carbonated beverages and colas) has increased, especially among children.[40] Additional sugars encompass a variety of sweeteners that may include sucrose, fructose, glucose, dextrose, maltose, and the high-fructose corn syrup (HFCS), which is used as a predominant sweetener in most sugar-sweetened beverages.[41] In 2000, SSB were the main source of energy from beverages consumed among Americans, contributing to 9.2% of the total energy intake,[42] a number more than two times higher than the 3.9% documented in the late 1970s.[43] On the other hand, the percent of energy intake from milk declined from 8.0% to 5.0%, with only minor changes in the consumption of all other beverages (i.e. alcohol, coffee and tea, fruit drinks and fruit juices).[43] Studies emphasize the negative health impacts of high intake of sugar,[44,45] and support the replacement of sugar-sweetened beverages with low-fat milk and 100% fruit juices.[46]

2.4.2 Nutrient-based Beverages

Nutrient-based beverages, such as milk and fruit juices, are also described as “foods that you drink” due to their ability to generate physiologic satiety, possibly from modifications in their protein and/or fibre contents.[47] The role of milk in human health has been well-documented.[48-51] Milk is a good source of vitamins and minerals (especially calcium), high quality fat, protein, and carbohydrate.[48] As well, it has protective effects against hypertension, cardiovascular diseases, cancers, and osteoporosis.[48-51] Recent studies in children and adolescents revealed a beverage consumption pattern with higher levels of sugar-sweetened beverages and lower intakes of nutrient-based beverages.[52-54] Similar displacement has been observed among American adults,[5,55] whereas the beverage intake pattern among Canadian adults is not yet fully understood, requiring further investigation.

In recent years, the consumption of 100% fruit juices has increased,[40] partially substituting the intake of raw fruit, however studies suggest that storage and packing methods for juices decrease the quality of this replacement.[56] Citrus fruit juices and cranberries are mainly consumed for their health benefits on cancers and arterial pressure,[57] and for prevention of bladder infection, respectively.[58]

2.4.3 Alcoholic Beverages

The average alcohol consumption (beer, wine, cocktails, spirits, and liquors) in the US has increased among moderate drinkers, and has decreased among heavy drinkers over the last five decades, with no decrease in the cumulative incidence of alcohol use disorders.[59] Beer is one of the top ten sources of energy intake among the US population.[42] The health impacts of low and moderate alcohol intake are controversial, however the excessive consumption of alcohol is associated with morbidity and mortality.[60]

2.4.4 Non-caloric Beverages

Non-caloric beverages include water, "diet" drinks sweetened with artificial sweeteners, and coffee and tea with no added sugar and/or cream. These beverages don't have any contribution to total energy intake from beverages.

The health impacts of adequate water consumption (metabolic and functional abnormalities prevention) have been well-documented.[61] Also, there is a relationship between high water consumption and healthier eating patterns which is more evident among higher-educated and older adults.[62]

Artificially sweetened beverages or diet soft drinks are consumed by less than one-fifth of American adults (17.5%).[5] Differently from the harmful effects of aspartame in animals,[63] there is no evidence that support its neurotoxic or carcinogenic effects in humans.[64]

Coffee and tea, which have been very popular among adults for thousands of years, often contain caffeine. Extensive research has been conducted to evaluate the health aspects of caffeine consumption. According to a review by Health Canada, moderate caffeine consumption (at dose levels of 400 mg/day or less) is probably not associated with adverse effects such as general toxicity, effects on calcium balance, and cardiovascular effects.[65]

2.5 Does Beverage Intake Affect Overweight and Obesity?

The relationship of beverage intake and overweight and obesity has been studied for many years. It has been concluded that an increase in both solid and liquid calories is associated with weight gain; however, the impact of liquid forms of energy intake on weight is stronger than solid calorie intake.[66-75] The impact of different beverage groups on weight are discussed and reviewed separately in the following sections.

2.5.1 Sugar-sweetened Beverages

Recently, many studies have been conducted to evaluate the link between consumption of sugar-sweetened beverages and weight gain.[66] There are four argued mechanisms that explain this potential association.[67] The simplest and strongest mechanism is based on energy balance where there is excess calorie intake from sugar solutions rather than solid preparations.[68-71] The other two mechanisms are related to the high glycemic load,[72,73] and low satiety[68,73] conferred by SSB. The last mechanism is explained by the displacement of milk from the diet, resulting in lower calcium intake, which could lead to weight gain since calcium is believed to have a role in reducing obesity.[75,76]

Reviewing five cross-sectional studies in adults,[41, 77-80] four studies[77-80] confirmed a significantly positive association between soft drink intake and body weight, however men had a less significant association than women. In the Healthy Worker Project, French et al.[77] found an independent association between soft drink consumption and higher body weight among female adults (n=1,913). Further supporting that result, Liebman et al.[78] showed that high intake of soft drinks could be an indicator for non-nutrient dense diets increasing the likelihood of weight gain. Also, in a study with adult females (n=2,419), Lin et al.[79] found positive association between percent of energy from soft drinks/juice drinks and BMI among women predominantly with higher socioeconomics status. Moreover, McCarthy et al.[80] concluded that in comparison to low calorie beverages, the intake of high calorie beverages such as soft drinks, squashes and cordials increased the likelihood of being obese by 3.9 times. Conversely, Sun and Empie[41] did not observe this relationship in the study of databases from the Continuing Survey of Food Intakes by

Individuals (CSFII)-1989–1991, CSFII-1994–1998, National Health and Nutrition Examination Survey (NHANES) III, and combined NHANES 1999–2002.

The results from seven prospective cohort studies show some evidence of positive association,[77,81-86] particularly in studies with a longer cohort period.[82-84,86] In a two-year study, Schulz et al.[81] reported higher intakes of soft drinks among male and female weight gainers than weight maintainers, however, when compared to female weight maintainers, women who lost weight did not report less consumption of soft drinks, a fact that might be explained by the study's design that classified mineral and tap water in the soft drinks category. Results from a 25.5-year study by Newby et al.[82] showed that a diet low in soda was associated with smaller gains in waist circumference and BMI. Moreover, in an eight-year cohort study among US healthy female nurses (n=51,603), Schulze et al.[83] reported that frequent intake of sugar-sweetened beverages may be associated with larger weight gain. Nooyens et al.[84] also found a positive association between increased sugar-sweetened beverage consumption and weight gain and waist circumference among 288 older adult healthy men who participated in a five-year cohort study. Bes-Rastrollo et al.[85] concluded that an increase in the consumption of sugar-sweetened beverages was associated with weight gain over a 28.5-month period. Finally, Dhingra et al.[86] reported an association between higher intakes of soft drinks and a 31% increase in the odds of developing obesity among 6,039 adults in a four-year follow-up.

Results from three short-term (≤12 weeks) clinical trials revealed that sugar-sweetened soda intake has a positive association with weight gain.[68,69,87] Moreover, a recent long-term (18 months) randomized, controlled, behavioural intervention trial study with more subjects (n=810) proved that a reduction in sugar-sweetened beverage intake will lead to weight loss.[88]

A recent meta-analysis[89] of 88 studies investigating the association between soft drink intake and health outcomes among children and adults showed that there is a clear association between soft drink consumption and increased body weight. Moreover, it was concluded that the differences in methodological factors (study design, gender, age, and beverage type) have substantial effects on study outcomes.[89] A summary of the literature on the association of sugar-sweetened beverages and weight in adults is provided in Table 2.2.

Table 2.2. Summary of literature on the association between sugar-sweetened beverages and weight in adults

Design	Authors (year)	Subjects (sex)	Age group(year) or Mean age (year)	Length of the study	Results
Cross-sectional	French et al.[77] (1994)	3,552 (F/M)	Mean ± SD= 38.1±10.3	N/A	Strong association between soda drinkers and BW in females
	Liebman et al.[78] (2003)	928 (M) 889 (F)	18-99	N/A	Positive association (stronger in females)
	Lin et al.[79] (2004)	2,419 (F)	≥20	N/A	SSB correlated positively with BMI in high-income samples
	McCarthy et al.[80] (2006)	1,379 (F/M)	18-64	N/A	Strong association between high-energy beverages and obesity
	Sun and Empie[41] (2007)	38,409 (M/F)	20-74	N/A	No relationship was found between obesity and SSB intake.
Cohort	French et al.[77] (1994)	3,552 (F/M)	Mean ± SD= 38.1±10.3	2 years	No significant association
	Schulz et al.[81] (2002)	17,369 (M/F)	24–69 (M) 19–70 (F)	2 years	In males: U shape association (both weight gainers and weight losers had higher intake of soft drinks compared with weight maintainers) In females: weight gainers had higher intake of soft drinks compared

					with weight maintainers
	Newby et al.[82] (2003)	240(M) 219(F)	Mean=60.8(M) Mean=57.3(F)	25.5 years	Positive association between soda and BMI and waist circumference
	Schulze et al.[83] (2004)	51,603 (F)	24-44	8 years	Higher intake of soft drink and fruit juice was associated with greater weight gain
	Nooyens et al.[84] (2005)	288 (M)	50-65	5 years	Positive association between SSB and weight gain
	Bes-Rastrollo et al.[85] (2006)	7,194 (M/F)	Mean=41	28.5 months	Limited association (only in those who gained 3-5 kg before study)
	Dhingra et al.[86] (2007)	6,039 (M/F)	Mean=52.9	4 years	Higher intake of soft drink increased odds of obesity among middle aged adults
Clinical Trial	Tordoff and Alleva[87](1990)	21 (M) 9 (F)	Mean=28.2 (F) Mean=22.9 (M)	9 weeks	SSB group showed increase in BW (stronger in Females)
	DiMeglio and Mattes[68](2000)	15 (M/F)	Mean=22.8	12 weeks	SSB group showed increase in BW
	Raben et al.[69] (2002)	6 (M) 35 (F)	20–50	10 weeks	SSB group showed increase in BW
	Chen et al.[88] (2009)	810 (F/M)	25–79	18 months	Reduction in SSB intake decreased BW

M=males; F=females; BW=body weight; SSB=sugar-sweetened beverages; BMI=body mass index.

2.5.2 Milk and Fruit Juices

Recent studies suggest a possible negative association between dairy product intake and body weight. Several cross-sectional studies in adults and children support this hypothesis,[90-94] however, such study designs are not strong enough to draw causal relationships. One possible mechanism to explain the negative association between dairy products and body weight is the calcium hypothesis [102,103] whereby calcium reduces intestinal fat absorption and increases fecal fatty acid excretion. Furthermore, it was concluded from a prospective randomized controlled trial on 14 overweight/obese women that both dairy and non-dairy sources of calcium with equal protein content promote statistically equivalent weigh loss. [104]

Data from cohort studies and randomized controlled trials do not show consistent results,[95-100] and the majority of the current evidence from clinical trials does not support the hypothesis that calcium or dairy consumption aids in weight or fat loss.[101] A different explanation for the weight-lowering effects of dairy products involves a mechanism that is calcium-independent and regulates fat metabolism by, relying on the high concentration of branched-chain amino acids in whey and angiotensin-converting enzyme (ACE) inhibitors that act as.[105-107] According to a recent experimental study on 45 young adult volunteers,[108] whole-fat dairy products significantly increased weight compared to low-fat dairy.

With regard to fruit juices, according to one study on children 2-11 years of age,[109] 100% juice intake was not associated with the likelihood of being overweight. More studies are required to evaluate this relationship among adults.

2.5.3 Alcoholic Beverages

Despite the fact that alcohol is a significant source of energy,[110] its effect on weight gain and obesity is still controversial among epidemiological studies.[111] It has been reported that in Japanese males aged 49 - 55 years, alcohol intake is positively and strongly associated with waist/hip ratio, but not with BMI.[112] Similarly, in a Swedish cohort study conducted with 70-year-old males, a significant positive association between high alcohol intake and abdominal obesity was found.[113] The effects of alcohol on body weight depend on different determinants of weight gain such as body composition,[114] gender,[115] physical activity,[116] dietary fat,[117] family history of overweight,[118] as well as the drink itself in terms of frequency and amount of consumption.[119] In population studies, the quantification of alcohol consumption poses challenges due to difficulties in identifying respondents with high intakes of alcohol and the possibility of underestimating alcohol intake among some consumers.[120]

2.5.4 Water, Coffee, and Diet Soft Drinks

There is evidence that drinking water may cause weight loss, and results from a long-term (one year) study showed that the absolute (≥1000 ml/day) and relative increases in water intake among 173 overweight women were significantly associated with weight loss.[121] Possible explanations for this include the replacement of caloric beverages with water, [122] and the increased water intake leading to metabolic changes .[121]

A recent review of epidemiological studies on the effects of coffee included a prospective study that assessed the effect of caffeine intake on weight change in American adults for over 8.4 years. The results showed a positive association between both caffeinated and decaffeinated coffee intake and weight loss.[123] A similar association has also been observed in a longer cohort study of twelve years.[124] However, the addition of sugar in coffee and tea might result on a different outcome.

With regard to artificially sweetened or diet drinks, results are controversial, but according to a systematic review by Malik et.al.,[8] the majority of studies reported neither weight gain after exposure nor actual weight loss. Conversely, a recent study documented a positive dose-response association between diet soft drink intake and long-term weight gain.[125] However, this discrepancy was attributed to the possibility of a pre-existing weight-gain pattern of individuals who switch to this drink in an attempt to lose weight.

2.5.5 Potential Confounders in the Association between Beverage Consumption and Overweight/Obesity

Studies suggest a relationship between obesity and different factors, such as age,[126,127] sex,[126-129] ethnicity,[130,131] rurality,[132-134] immigration,[135,136] marital status,[128,129,137,138] education,[126,133,139] income,[126,140-142] physical activity,[139,143,144] as well as smoking habits.[145,146]

Likewise, these variables are also proved to influence the frequent consumption of beverages,[147-149]including sugar-sweetened beverages.[150,151] In addition to such variables, the accuracy of surveys can be further improved when seasonal changes are taken into consideration.[152] Hence, the inclusion of all potential confounders in the relationship between beverage intake and obesity can provide less biased and more precise results.

2.6 Cluster Analysis and Dietary Pattern Assessment

Due to the complexity of dietary behaviours, a variety of methods to assess such patterns is required. As discussed earlier, the Food Frequency Questionnaires (FFQ) and the 24-hour dietary recalls are the most commonly used methods to assess usual dietary intake.[16] Some methods of dietary pattern measurement include Principle Component Analysis (PCA), Cluster Analysis (CA); and most recently, Reduced Rank Regression (RRR).[153] The key goal of a dietary pattern assessment is to investigate its association with health outcomes.[153] The first two methods, PCA and CA are used to describe actual intake patterns in the population, and these methods have good reproducibility. However, RRR defines a linear combination of food intakes, and is less reproducible because it is based on relationships with outcomes rather than consistent behavioural patterns.[153]

Unlike the PCA method, which develops patterns based on relationships between food groups, cluster analysis derives dietary patterns based on differences in intakes.[154] Cluster analysis is used to identify homogeneous groups in a population, thus minimizing within-group variation and maximizing between-group variation.[155] CA can also define distinct patterns of dietary intake, and provide a clear description of patterns in the population by presenting mean values for food groups within each cluster.[153]

One of the limitations of cluster analysis is the introduction of subjectivity which is plausible at various points.[153] One of the challenges is that the grouping of dietary items can be done by using prior knowledge of dietary patterns in a population; and another is the selection of the type of cluster analysis procedure. One of the four general approaches to cluster analysis is the “K-means method”, which is appropriate for large populations. The “K-means method” (FASTCLUS procedure in SAS, 2009) categorizes participants into non-overlapping groups in an iterative process by comparing Euclidean distances, but the researcher needs to pre-specify the number of clusters.[155] Finally, there is the selection of a final pattern or the best solution, which can be resolved by a number of criteria such as plotting of the magnitude of factor loading, exploring multiple solutions, and most commonly, confirming the interpretability of the results.[154,155]

Because the increased interest in applying cluster analysis to assess food intake patterns has been relatively recent,[153-158] just a few studies have attempted to use this method when investigating beverage consumption patterns.[5,62,82,155,156] Moreover, earlier reports confirmed the usefulness of this method in epidemiologic studies that investigated obesity as an outcome.[157,158] To our knowledge, only two studies, those by Newby et al.[82] and by LaRowe et al.,[155] applied cluster analysis to evaluate the relationship between beverage intake and BMI. Schulz et al.[81] recommended using cluster analysis for further exploration of this relationship, and McCarthy et al.[80] explained that they did not employ this method because the groups in their study were predefined based on BMI.

Chapter 3

METHODOLOGY

To address the study objectives, we used secondary data from the Canadian Community Health Survey (CCHS) Cycle 2.2. In this chapter, we will provide detailed information on this survey and describe the analytical approach used to assess the data.

3.1 Canadian Community Health Survey 2.2

The Canadian Community Health Survey (CCHS) Cycle 2.2 is the most recent and comprehensive source of information on the dietary intake of Canadians. The CCHS 2.2 data was collected between January and December of 2004 through a nationally representative cross-sectional survey, in response to an essential need for more timely and reliable information about the nutritional status of Canadians.[159] The CCHS 2.2 is composed of two different parts: (1) General Health component and (2) 24-hour dietary recall. The general health component includes information on selected health conditions, socio-economic and demographic characteristics of 35,107 Canadians, as well as their measured height and weight.[159] Hence, this survey provides a more accurate measurement of overweight and obesity compared to self-reported height and weight which tends to underestime weight and overestimate height leading to lower prevalence of overweight and obesity.[160]

The 24-hour dietary recall component provides information on the types and amounts of food and beverage, as well as on the nutrient intakes during the 24 hours prior to the interview.[159] In addition, a second recall involving almost 30% of the participants was conducted on a different day of the week, 3 to 10 days after the first interview. The 24-hour dietary recall component is subdivided into two files: Food and Ingredients Details (FID) and Food Recipe Level (FRL). These files are similarly structured based on a unique sample identification number. For each participant, the FID file contains the nutrient values for basic food and ingredient food levels, while the FRL file contains nutrient values for main recipe and sub-recipe levels. A separate file labelled FDC (Food Description) contains descriptive information on foods that are used in the coding and grouping of data into the FID and FRL files.[159] The FDC file does not contain identification numbers, but does contain food groupings and descriptions, in addition to food codes that were available to the CCHS 2.2 during the coding of the reported foods (2004). These food codes were based on the Nutrition Survey System (NSS), the Canadian Nutrient File (CNF), and the Bureau of Nutritional Sciences (BNS) at Health Canada. The CNF is continually updated and contains nutrient values that are largely based on information from the United States Department of Agriculture Nutrient Database for Standard Reference, but with Canadian regulations, fortification values, and unique Canadian foods (CCHS 2.2 User Guide, 2008).[159] The data files of CCHS 2.2 are depicted in Figure 3.1.

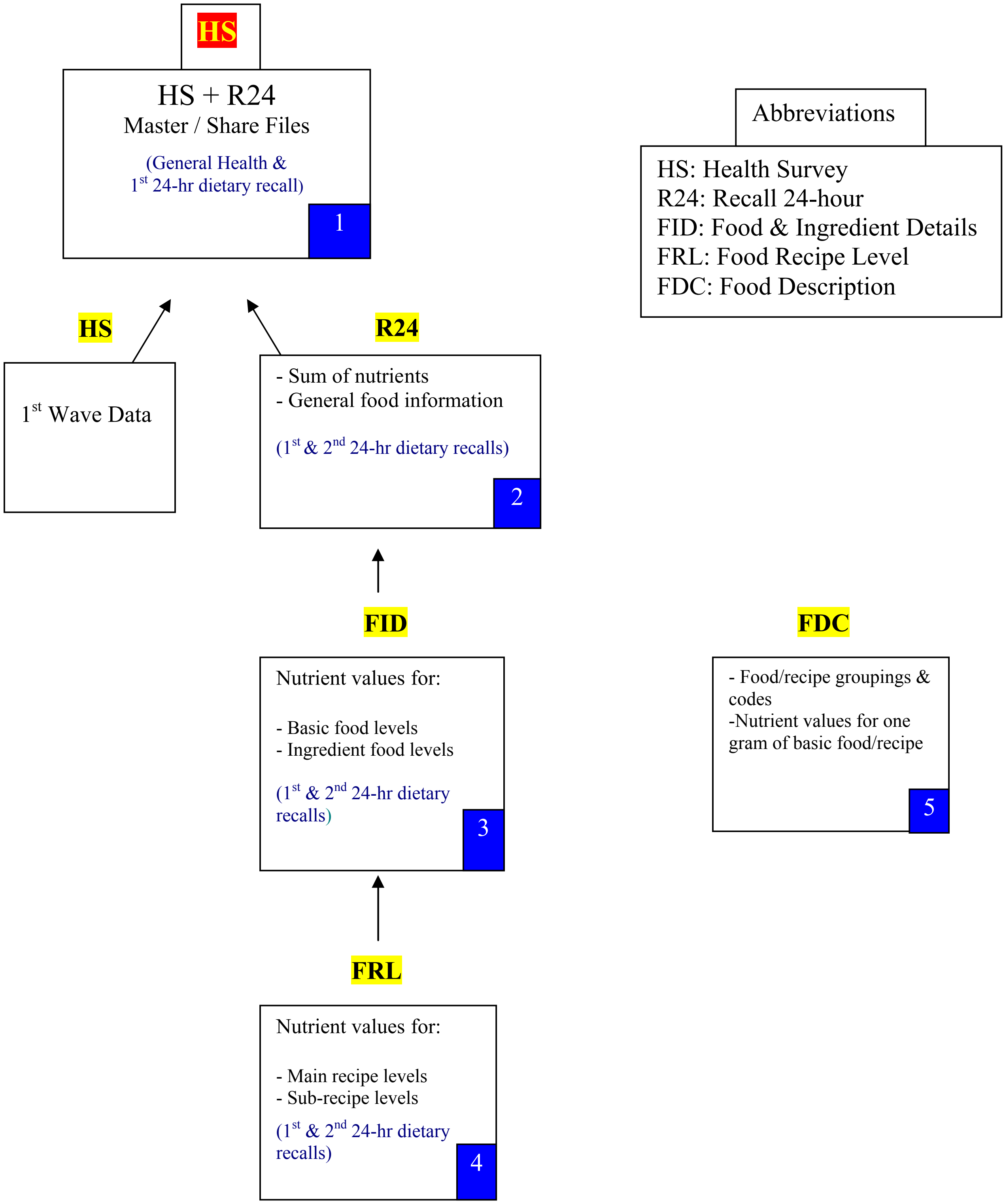


Figure 3.1. CCHS 2.2 Data Files

3.1.1 Sample Design

Statistics Canada used a complex stratified cluster design to obtain a representative sample of Canadians in terms of age, sex, geography, and socioeconomic status.[159] The target population included residents of all ten provinces, but excluded persons living in the three territories, Indian Reserves, Institutions, and Canadian Forces. To assign the sample to the provinces, a two-step strategy was used. First, in each province, the combination of person-level sampling and two-framed household-level sampling strategy provided 80 sample units per each 14 age/sex groups (1-3 all, 4-8 all, 9-13 M/F, 14-18 M/F, 19-30 M/F, 31-50 M/F, 51-70 M/F, 71year or more M/F).[159]

In the second stage, the remaining units were distributed to the provinces by a power-allocation method. Moreover, the provinces of Ontario, Manitoba, and Prince Edward Island paid for larger samples prior to the start of the data collection. The purpose of these buy-ins in Manitoba and Ontario was to provide reliable estimates for sub-provincial areas, and in Prince Edward Island, to increase respondents in some age/sex groups in order to get satisfactory sample sizes. Then, within provinces, the sample was assigned into two urban and rural strata. Also, a larger sample of off-reserve aboriginals was included in the sampling to obtain national-level distributions. Therefore, this multistage survey design provided a representative sample of dietary intakes at the national and provincial levels.[159]

3.1.2 Data Collection

In 2003, a 700-unit field test was conducted to examine the questionnaire. Interviewers received training for 3.5 days, and any problems encountered during data collection were reported to a senior interviewer.[159] Data were collected from January 14, 2004 to January 21, 2005, and a computer-assisted interviewing (CAI) technique was used to complete the questionnaires. The first interview (60 minutes) was conducted in person in the respondent's home and the second recall (30 minutes) was completed over the phone. No proxy interview was conducted for individuals above 12 years of age. Weight was measured with portable high quality scales, and for standardization purposes, interviewers were appropriately trained and monitored to measure height.[159]

3.1.3 Data Processing

Most of the editing was completed by using the CAI technique during the interviews. Successful completion of the questionnaire's minimum requirements was defined as a "partial interview", and these included the complete 24-hour dietary recall component, as well as a minimum part of the general health questionnaire, up to the end of the Food Security module. Anything that provided less information than a "partial interview" was classified as a "non-response".[159]

Open-ended questions were examined and coded later at the Head Office. A number of derived and grouped variables were created to facilitate the data analyses. To derive meaningful estimation from the survey, the weighting step was performed. The confidentiality was assured by Statistics Canada, and the survey "master file" was protected from public-use access.[159]

3.1.4 Weighting

The principle behind "weight" adjustments involves the estimation that each participant in a probability sample represents a number of people that are outside the sample. For example, in a simple random 4% sample (1/25) of the population, each person in the sample has a "weight" of 25, or represents 25 people in the population.

In the weighting step, two "weight" variables, WTSD_M and WTSD_MHW, were computed and developed by treating the five sampling frames individually to the integration step. Then some adjustments were carried out to create the final weights. The CCHS 2.2 guidelines recommend that the master weight (WTSD_M) that appears for each record must be applied to obtain meaningful estimates from the survey.[159] The second weight variable (WTSD_MHW) was created due to a high non-response rate of the measured weight and height, and is specifically used when height and weight are included in the analysis.[159]

3.1.5 Data Quality

The target population, sample size, and allocation were appropriately designed to meet the CCHS 2.2 objectives.[159] Also, a second 24-recall was collected to minimize within-person variation and to provide reliable information about the dietary intake and nutritional status among the representative sample.[159] However, the monitoring of a large group can be met by using a single 24-hour recall or record if all the days of the week are equally represented.[15]

Moreover, the application of CAI in data collection offered many quality advantages above other collection methods such as automatic customised question texts and immediate feedback to the interviewer when an invalid entry is recorded.[159] Additionally, known for its well-established validity, the Automated Multiple-Pass Method with a five-step approach was applied to help subjects to recall food and beverage intake.[17-19] Finally, the satisfactory national response rate of 76.5% was achieved in CCHS 2.2. The possibility of both sampling and non-sampling errors is indicated in the Statistics Canada User Guide Catalogue.[159]

3.1.6 Guidelines for Statistical Analysis in CCHS 2.2, Weighting and Bootstrapping

Analysts face challenges when using data from a complex survey design such as the CCHS 2.2. The unequal probabilities of selection, as well as the stratification and clustering pose difficulties in the estimation and calculation of variance.[159] Unequal probabilities of selection can be resolved by the application of "weight" adjustments. However, the stratification and clustering are not taken into account by weighting. Statistics Canada has produced a set of tables of Coefficients of Variation (CV) to provide a means of assessing the quality of tabulated estimates; however CV tables provide only approximate values and are unofficial.[159] There is no simple mathematical formula to calculate the exact coefficients of variation that would fit into the entire CCHS sampling frame; therefore, CCHS 2.2 recommends using the bootstrap method for resampling to estimate measures of precision at the population level.[159]

3.2 Analytical Approach

Data analyses were conducted in three stages: 1) cleaning, processing, and descriptive analysis; 2) determining the pattern of beverage intake; 3) advanced statistical modeling.

3.2.1 Data Cleaning, Processing and Descriptive Analysis of Beverages

This stage comprised the necessary steps for data cleaning and processing, creating new variables, aggregating beverage and nutrient data as well as demographic, anthropometric, and socioeconomic data from different files, and creating the final data set for the next stages. This stage also included defining inclusion and exclusion criteria.

3.2.1.1 Data Cleaning and Processing

The following steps were conducted through a multiple-step procedure of detection, classification, coding, merging and correction. Beverages were identified and isolated from the FDC file.

Identification

This was completed using three variables within the FDC file: i) FIDD_CDE, which contained unique food codes for 8,784 food items (5,246 basic food levels and 3,538 main recipe levels) corresponding to the codes in the NSS; ii) FDCD_DEN, which provided the names (in English) of food items according to CNF; and iii) FDCD_FGE, showing the food group descriptions (in English) from the BNS (Appendix B, flow charts 1.a and 1.b).

Classification

Following the identification of beverages, the FDCD_DEN and FDCD_FGE variables were used to classify beverages. The information provided by these variables allowed for classification of beverages based on their energy and nutrient content. Information in the FIDD_CDE variable was used to search the CNF for some beverages to assist with appropriate classification. Based on energy and nutrient content, beverages were first categorized into 17 categories (Table 3.1). Further, categories were merged into 4 groups: sugar-sweetened low nutrient beverages, nutrient-based beverages, alcoholic beverages, and beverages with no additional energy (Table 3.1). Beverages used in food recipes of non-beverage foods category (For instance, water, milk and wine in a cake or food recipe) were not considered as beverage. Condensed milk and milk added to tea, coffee or cereals were excluded. Undiluted beverages and any beverage with less than 35% moisture were excluded. Detailed steps of this stage of analysis are summarized in Appendix B (flow charts 1.a and 1.b).

Table 3.1. Beverage classification and definitions identified for CCHS 2.2

Beverage Category	Types of beverages in category
	I-Sugar-sweetened Low Nutrient Beverages
1- Fruit Drinks	Fruit drink <100% fruit juice from concentrate, fruit flavoured beverages, lemonade, fortified or not
2- Regular soft drinks	Carbonated beverages, colas, clear sodas, fruit-flavoured sodas (eg lemon-lime), non-alcoholic beer
3- Tea, sweetened	Iced tea, spiced, instant
4- Coffee, sweetened	Instant, brewed, flavoured (eg. cappuccino, mocha, but with whitener not milk)
	II- Nutrient-based Beverages
5- Plain milk	Skim, 1%, 2%, whole milk, and those with added milk solids, diluted evaporated milk, and extra calcium
6- Milk-based beverages >2%MF	Milk shakes, iced cappuccino, eggnog, whole chocolate milk, hot chocolate with whole milk, malted milks- unless specified with low fat milk, coffee substitute mixed with whole milk, milk-based smoothie, latte
7- Milk based beverages: low-fat ≤ 2%	Skim milk latte, low fat chocolate milk (or where MF was not specified), mixed milk beverages where low-fat milk was specified (eg. eggnog mix, instant breakfast)
8- Other types of Milk	Soy based beverages, goats milk, rice beverage, whey beverages, buttermilk, sheep milk, protein shakes
9- Vegetable juice	100% vegetable juices (carrot, tomato)
10- Fruit juice	100% fruit juices, sweetened fruit juice
	III- Alcoholic Beverages
11- Beer	All types of beer
12- Wine	All types of wine
13- Other alcoholic	Cocktails, spirits, liquors, coolers
	IV- Non-Caloric Beverages
14- Water	Tap and bottled water, club soda, Perrier type
15- Diet drinks	Colas and non-colas with sugar-substitute (aspartame), fruit drinks with aspartame or sucralose, "low-calorie", non-alcoholic wine
16- Tea	No added sugar/cream; brewed, instant, herbal
17- Coffee	No added sugar/cream; brewed, instant, caffeinated, decaffeinated, chicory

Coding

Based on the beverage categories that have been created, all the identified beverages were assigned a code (1 through 17). To do this, a beverage category variable was created using SAS software (version 9). The beverage variable categorized over 400 unique beverage codes (FIDD_CDE) into the 17 specified beverage categories. The new beverage category variable was exported to the FID and FRL files. These files contain dietary information from the initial 24-hour recall and the second 24-hour recall, which was completed with 30% of the initial sample to determine usual intake. Referring to the CCHS guidelines[19], and Figure 2.1, only the information from the initial 24-hour recall were used for this study.

Merging

After removing data from the second recall, the SPSS program was used to merge the FID and FRL files into one file containing all the beverage data (SAS). Subsequently, data from the newly created file were evaluated to avoid repetition. The structure of the merged FID/FRL file is such that for each individual surveyed (n=35,107), there are multiple rows of dietary information (1,299,994). These multiple rows represent various beverage items consumed by the individual in the 24 hours prior to the interview. As an example, if an individual had 13 intakes of six different kinds of beverages during the period in question, 13 lines of information (rows of a spreadsheet) would be present for that participant. In order to have a compatable file with HS file, we needed to calculate total gram/kcal intake of different beverages for each participant. By using the SORT command in SAS (Appendix C), the format of the merged FID/FRL file was converted so that each partcipant had only one line of data, essentially summing the volume/energy of beverages consumed. In this step the data from the merged FID/FRL file after changing the format was merged with the HS file.

Correction

At this stage, missing information on beverage intake of participants was identified and recoded as zero (n=82). As well, there were some missing values of beverage intake coded by CCHS as 100,000 which needed to be changed to zero to avoid miscalculation. The decision to recode missing dietary information to zero was based on previous studies showing that participants who did not indicate the consumed amount of a specific food item were most likely to have left the item blank because they did not consume this product.[161,162] The resulting file was saved and labeled as a beverage file. The beverage file was then merged with the HS file according to the unique individual identification numbers. This created a file that contained demographic, anthropometric, and socioeconomic variables, as well as beverage intake information (volume in gram/energy in kcal) from over 35,000 subjects in the CCHS 2.2.

Descriptive Analysis

This was performed by calculating the average daily consumption of defined beverages (in grams) among Canadian adults. In this initial step, 14,304 subjects were divided into six age/sex groups (19-30 year, 31-50 year, and 51-65 year). Further, one-way ANOVA was employed to assess meaningful differences among different age/sex groups. Then, the main groups of caloric beverages (Sugar-sweetened, Milk categories, Juice categories, and Alcoholic beverages) were explored. Four groups of milk, regardless of fat additives, were merged into one group named "milk"; and the fruit juice and vegetable juice categories were merged into a single group called "juice".

Following the same approach used to calculate the average daily intake of beverages in grams, in this step, the energy intake of different beverage groups were calculated for adult participants among six age/sex groups. Data were weighted to take into consideration the inverse probability of being in the sample, and bootstrapped to account for the complex survey design. This stage of data analysis was conducted using the STATA SE 10 software in the SKY Research Data Center at the University of Saskatchewan.

3.2.1.2 Demographic, Anthropometric, and Socioeconomic Variables

Variables that the literature recognizes to influence the relationship between beverage intake and BMI will be covered in this section. In addition, the classification of BMI for adults, as well as the processes involved in creating new variables, aggregating data for age/sex groups, and recoding categorical variables for household income, education, ethnicity and household food security status will be described here.

The "Age" variable indicates the age of respondents in the CCHS 2.2. Two forms of variables, categorical (DHHDDDRI) and continuous (DHHD_AGE), were employed in the analyses. Based on Dietary Reference Intakes (DRI),[19] the categorical age/sex variable categorized the respondents into 14 different sex/age groups. This variable was used in the descriptive analysis of the beverage intakes of Canadian adults. The continuous form of the age variable was applied in the final modeling analysis. The “Respondents’ Residence” variable was divided into two areas classified as “rural” and “urban”. The “Immigration Status” variable also divided subjects into two groups: the “immigrant” and the “non-immigrant” groups.

The “Ethnicity” variable indicates the ethnic, racial or cultural origin of the respondents. Ethnicity or racial background was self-identified and included 13 options for respondents to select, including “Other”. Initially we were interested to categorize subjects into three groups, but due to small cell sizes in some age/sex groups, ethnicity was ultimately merged into two categories (Table 3.2).

Table 3.2. Original and merged ethnicity categories

Original ethnicity categories	Initial merged ethnicity categories	Final merged ethnicity categories
White	White	White
Aboriginal	Aboriginal	Non-white
Southeast Asian	Other	
Korean		
Filipino		
Japanese		
Chinese		
Black		
South Asian		
Arab		
West Asian		
Latin American		
Multiple		
Other		

Household income was recorded as total income of all household members before taxes and deductions from all sources in the last 12 months. CCHS 2.2 introduces various derived variables on income with different classifications. To determine the income levels, we used the “Income” variable with five categories created by CCHS 2.2, which was based on the combination between income and the number of people living in the household (Table 3.3).

Table 3.3. Household income levels in CCHS 2.2

Categories	Total household income and number of people
Lowest	<$10,000 if 1-4 people, or <$15,000 if 5+ people
Lowest middle	$10,000-14,999 if 1-2; $10,000-19,999 if 3-4; $15,000-29,999 if 5+ people
Middle	$15,000-29,999 if 1-2; $20,000-39,999 if 3-4; $30,000-59,999 if 5+ people
Upper middle	$30,000-59,999 if 1-2; $40,000-79,999 if 3-4; $60,000to 79,999 if 5+ people
Highest	≥$60,000 if 1-2 people, or ≥$80,000 if 3+ people

The "Physical Activity Index" variable (PACDDEE) was derived from Daily Energy Expenditure values (kcal/kg/day). This variable was created to measure the respondents' average daily energy expended during leisure time activities in the past three months. Based on the Ontario Health Survey (OHS), the respondents of age 12 and older in the CCHS were categorized into three groups (Table 3.4).

Table 3.4. Physical activity index categories in CCHS 2.2

Physical activity index	Daily energy expenditure
Active	$3 \leq$ PACDDEE < NA
Moderate	$1.5 \leq$ PACDDEE < 3.0
Inactive	$0 \leq$ PACDDEE < 1.5

There was not a Season variable in the CCHS 2.2, and to create this variable, the distribution of participants throughout the interview months was carefully examined. Moreover, the seasonality of beverage intakes among Canadian adults was also investigated for those months. Based on these findings, as well as on the weather conditions in Canada, the variable "Season" was created and classified into three categories (Table 3.5).

Table 3.5. Season variable categories created in this study

Season	Months
Winter	January, February, March, November and December
Spring/Fall	April, May, September and October
Summer	June, July and August

Based on the smoking status, the “Smoking” variable specifies the type of smoker that the respondent is. Originally, the variable had 6 categories, but these were later merged into three groups to avoid small cell sizes (Table 3.6). Unlike in previous cycles, in this survey, respondents were not asked if they had ever smoked a whole cigarette.

Table 3.6. Smoking status categories

New categories	CCHS 2.2 categories
Current smoker	Daily smoker Occasional smoker that was a former daily smoker Occasional smoker
Former smoker	Former daily smoker Former occasional smoker
Never smoker	Never smoker

Income-related household food security information, adopted from the United States Food Security Survey Module, was available in the CCHS 2.2 and was collected using the Household Food Security Survey Model (HFSSM). This Model is an 18 item questionnaire that “focuses on self-reports of uncertain, insufficient or inadequate food access, availability and utilization due to limited financial resources, and the compromised eating patterns and food consumption that may result.” (Health Canada, 2007b). The tool measured food security for adults and children in the household, and based on the responses given, the household was assigned to a particular level of food security: food secure, food insecure - moderate (without hunger), and food insecure - severe (with hunger). Due to small cell sizes, we merged moderate and severe food insecurity levels, and recoded the “Food Security Status” variable into two categories: food secure and food insecure.

The “Respondent Education” variable indicates the highest level of education obtained by the respondent. Two categories from a total of four were merged to avoid small sample sizes in the cells. Therefore the new variable has three categories (Table 3.7).

Table 3.7. The respondent education classification

New categories	CCHS 2.2 categories
<Secondary school	Less than secondary school graduation
Secondary graduation	Secondary school graduation
Post-secondary degree	Some post secondary education Post secondary graduation

The "Marital Status" variable categorizes the respondents into 6 groups according to their marital status, and these include: married, common-law, widowed, divorced, separated, and single or never married". To avoid small numbers in the cells, some groups were merged. Therefore, the new variable was coded into three groups: 1- married or common-law, 2- widowed, divorced, or separated, and 3- single or never-married.

The energy intake from the "Food" variable is the total intake (kcal/d) from all food and beverage sources within 24 hours prior to the interview (excluding vitamin and mineral supplements). This variable is presented in kilocalories, one kilocalorie being the amount of energy needed to increase the temperature of one kilogram of water by one degree Celsius at one atmosphere pressure.

The CCHS 2.2 provides measured height and weight for 63% of participants (as well as self-reported values). To minimize non-response bias[160], a specific weighting value was created by Statistics Canada. In our study, data on measured height and weight, in addition to specific weighting variable were used that allowed the measured data to be generalizable at population level. In the CCHS 2.2, the BMI variable had two forms: the continuous (MHWDDBMI) and the categorical (MHWDGISW). For our study, a new categorical BMI variable was created based on the WHO and Health Canada classification (Table 2.1). The portion of the sample classified as underweight was very small (1%), being irrelevant to our study and therefore, excluded from the analyses. Three categories of obesity were merged into a single obesity group, resulting in three categories of "BMI" variable: normal, overweight, and obesity.

3.2.1.3 Inclusion Criteria

For this study, individuals above 65 or below 19 years of age, as well as pregnant and breastfeeding women were excluded. This resulted in our first population of interest with a final sample size of 14,304 (Figure 3.2). Moreover, because under-reporting of energy intake can occur in studies examining the association between diet and BMI,[163] we decided to identify those who reported implausible energy intakes. Such screening was possible by using the equations from the Institute of Medicine (IOM) (Table 3.8) to calculate the respondents' basal predicted energy expenditure.[164]

Table 3.8. Equations for BMR calculation by IOM[164]

Sex	Age	Equation
Women	18-30 Year	(14.7 × W)+496 ±121
	30-60 Year	(8.7 × W) +829±108
	>60 Year	(10.5 × W) +596±108
Men	18-30 Year	(15.3 × W) + 679±151
	30-60 Year	(11.6 × W) +879±164
	>60 Year	(13.5 × W) +487±148

W, weight in kilograms

The IOM equations require the respondents' weight information to calculate their Basal Metabolic Rate (BMR). In the CCHS 2.2, almost half of the survey population had their weight measured, while some other individuals only had self-reported weight information. For those who had neither measured nor self-reported weight information, weight was estimated by the mean weight calculated in their corresponding age/sex groups. Energy intakes corresponding to less than 70% of the respondents' predicted basal energy expenditure were considered implausible[163] (n=2,478), as well as those who had a missing value for energy intake (n=15). A new variable "Plausible Energy Intake" was created with a final sample of 11,811 subjects, representing our second population of interest. For details, see Figure 3.2.

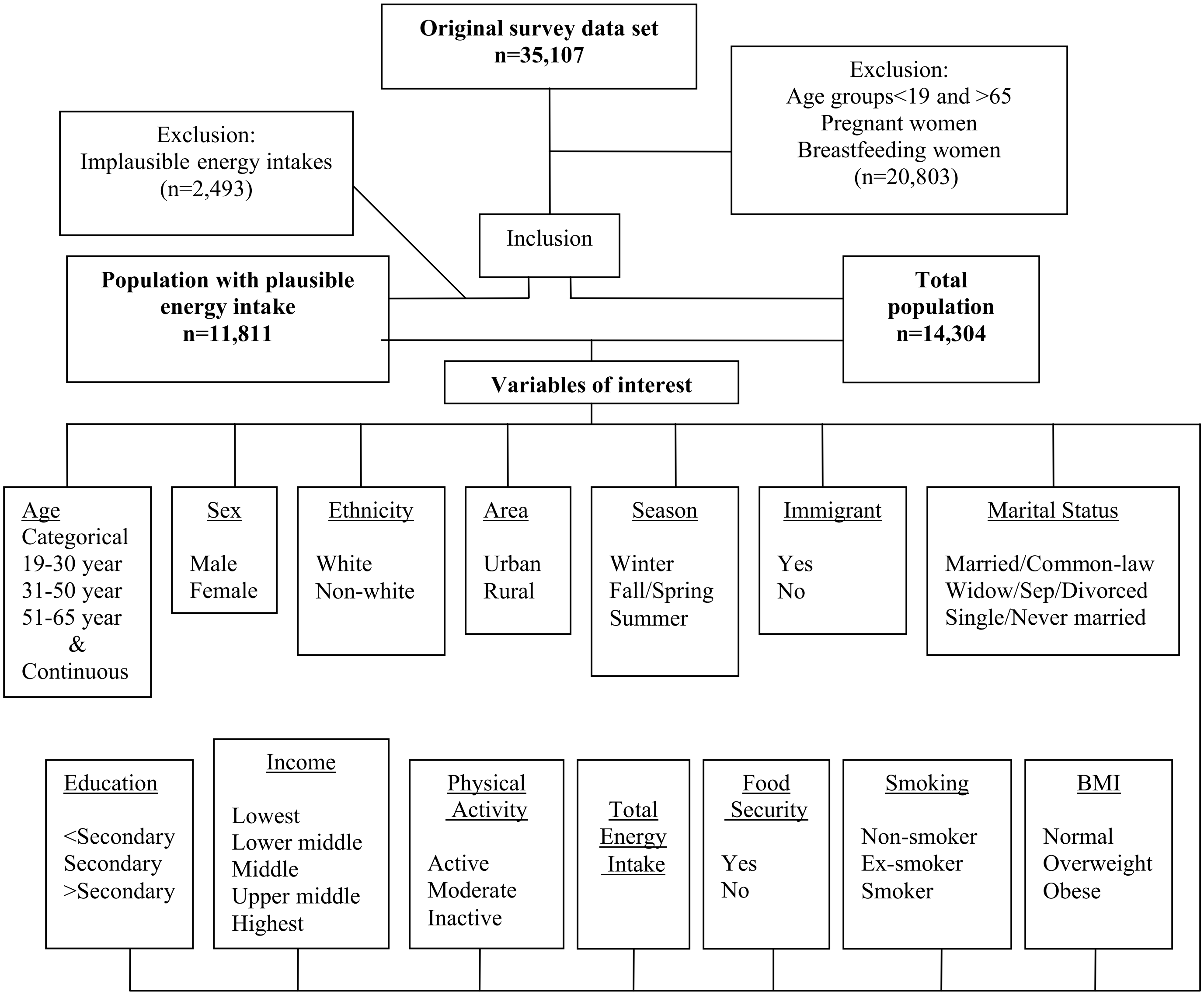


Figure 3.2. Inclusion criteria and variables of interest in the final analysis

3.2.2 Determining the Patterns of Beverage Intake

In this study, Canadian adults were classified into distinct groups based on the predominant patterns of beverage intake. Following a similar procedure as LaRowe at al.,[155] we used a cluster analysis K-means method to determine the patterns of beverage intake. With the K-means method, it is necessary to predefine the number of clusters - a procedure that is sensitive to outliers.[155,165] In order to take such outliers into account, subjects whose intake of a beverage group was ≥ 5 standard deviations from the mean were temporarily removed. We initially started the procedure with 40 clusters (a number higher than two times our 17 beverage categories), and temporarily removed subjects in clusters with less than or equal to 5 members. Once the optimal number of clusters was determined, all the temporarily removed subjects were included back into the dataset.

Optimal number of clusters was determined using Cubic Clustering Criterion (CCC), Pseudo-F statistic (PFS), and interpretability.[164] The values of CCC and PFS for each cluster set with no outliers (3 to 20 cluster sets) were recorded and then assessed using a line graph to identify 'local peaks' for the values.[164] The cluster set(s) at which CCC and PFS both peaked were explored for interpretability. Interpretability was assessed using the mean intake of each beverage category within each individual cluster. For example, where mean milk intake among individuals in Cluster 1 was significantly higher than among any other clusters, this cluster was identified as a unique cluster and named "milk cluster". For interpretability to be achieved, a clear and unique pattern of intake for each cluster was necessary. The steps taken for cluster analysis are summarized in Appendix B (Flow chart 3, 4). From all cluster set analyses, the cluster(s) with the highest number of participants emerged as a mixed cluster(s), where no dominant pattern of beverage intake was (were) found. This observation was consistent with other cluster analyses on dietary intake representing the majority of the population who consume a variety of beverages in their diet.[155] After reviewing the preliminary results, we decided to conduct more comprehensive analyses to obtain an optimal distribution of participants in the clusters, taking into account factors such as age and sex.

Once the optimal number of clusters was identified, the mean intake of each beverage group (g/d), total beverage consumption (g/d), and the mean percentage of total energy intake from beverages were calculated in all clusters. The means and frequencies for sample characteristics (e.g., household education, food security, ethnicity, and income) were determined. Finally, to test differences across beverage patterns, general linear models were used (one-way ANOVA test for continuous variables, and χ^2 test for categorical variables). Alpha was set at the 0.05 level. All analyses were first performed un-weighted, later going through the weighting and bootstrapping steps.

3.2.3 Statistical Modeling

In this step, we conducted regression analyses to evaluate the association of patterns of beverage consumption in Canadian adults with overweight, obesity and overweight/obesity (defined by BMI), while controlling for potential confounders. As the initial step, univariate logistic regression analyses were conducted to examine the association of individual potential confounders (i.e. age, area of residence, season, ethnicity, marital status, immigration status, smoking habits, food security, energy intake from food, physical activity index , respondent education level, and household income) with outcome variables (i.e. overweight, obesity and overweight/obesity).

Independent variables with a significant association ($p<0.2$) and biological justification with overweight/obesity were included in the multivariate model of binary logistic regression analyses, STEPWISE procedure.[166] For variables with collinearity (e.g. income-related household food security and household income), the variable with stronger relationship to overweight/obese was included in the full model. These steps were carried out for both sex groups in the total population, and in the population with plausible energy intake. Following recommendations from Statistics Canada, all regression analyses were weighted and bootstrapped using measured height-weight weighting variables to obtain estimates at the population level. Results are presented as odds ratios (OR) with 95% confidence intervals (CI). In all analyses alpha was set at the 0.05 level, unless otherwise mentioned. A variety of statistical packages were used in different stages of analyses, including SAS, STATA, SPSS, and EXCEL.

Chapter 4

RESULTS

In this chapter, the results of the study are presented in three stages including i) descriptive analysis of beverage intake among Canadian adults (19-65 years) by six age/sex groups; ii) determining patterns of beverage intake among Canadian adults using cluster analyses; and iii) advanced statistical analysis which includes results from linear and logistic regression analysis.

4.1 Descriptive Analysis of Beverages

To assess and compare beverage intake of Canadian adults among different age/sex groups, total gram intake of beverages and total energy intake from each beverage were calculated among 14,304 Canadian adults: 6,825 males and 7,479 females by three age groups (19-30 year, 31-50 year, and 51-65 year). All results were weighted and bootstrapped using appropriate weight variables.

4.1.1 Descriptive Beverage Intake among Age/Sex Groups

The estimated mean daily intakes of beverages (gram) and standard error of the mean (SEM) in 17 beverage groups and four main beverage categories in all Canadian adults are presented in Table 4.1. Total beverage intake of men was significantly higher than women in all three age groups ($p<0.05$). Top three beverages consumed by males and females were water, coffee, and soft drinks. Men drank significantly more sugar-sweetened beverages and more alcoholic beverages than women in all age groups ($p<0.05$). In men, total beverage intake was the highest in younger adults (aged 19 to 30 years), and the lowest in 51-65 years old men. Daily beverage consumption of 19 to 30 year old men averaged 2,755 grams/day and 2,180 grams/day for 51-65 years old men. This reduction was observed in the three main groups of beverages excluding non-caloric beverages in which no significant change was observed across age groups. The top three choices of beverages among younger men (aged 19 to 30 years) were water (1114 grams/day), soft drinks (303 grams/day), and beer (298 grams/day). In older men (31-50 years and 51-65 years), the top three choices were water, coffee, and beer (Table 4.1).

Water was the top beverage consumed by all age groups in men, although its consumption decreased in older men. Beer and soft drinks were consumed at the similar amounts (298 and 303 grams/day, respectively) by younger men and again the consumption of beer and soft drinks were lower in older age groups. Conversely, the consumption of coffee increased in older adults (458 grams/day). Plain milk, fruit juice, and fruit drink were the second top three beverages consumed by men mostly by younger adults (Table 4.1).

In women, total beverage intake was the lowest among older adults (1,987 grams/day). There was no significant change in total beverage consumption by 19 to 30 year old and by 31-50 year old women (2,186 and 2,277 grams/day). The consumption of non-caloric beverages was the lowest among younger female adults (1,365 grams/day) and the highest among women aged 31 to 50 years (1,677 grams/day). The consumption of sugar-sweetened and nutrient-based beverages among older women had lower amounts compared to younger women. There was no significant change in consumption of alcoholic beverages among different age groups in women. The top three choices of beverages among women aged 19 to 30 were water, milk and coffee; and in other age groups of women water, coffee, tea and milk were consumed the most (Table 4.1).

Similar to men, water was the top beverage consumed by all age groups of women, and the consumption was the lowest in older women adults (887 grams/day). The consumption of coffee was the lowest in younger adults (179 grams/day). The intake of tea was higher in older age groups of women compared to younger women; yet plain milk consumption didn’t change significantly across age groups. Soft drinks, fruit juice and fruit drink were the second top three beverages consumed by women mostly by younger adults (Table 4.1).

Beverage intakes of Canadian adults in four main beverage categories and total beverage consumption are illustrated on Figures 4.1 to 4.5.

Table 4.1. Mean ± standard error of the mean (SEM) daily intake (grams) of beverages among different age/sex groups

Beverage categories	19-30 years		31-50 years		51-65 years	
	Men	Women	Men	Women	Men	Women
Fruit drink	156.6±12.4[e]	126.2±11.3*[e]	102.9±9.2[e]	67.4±6*	50.4±5.3[e]	46.8±4.6
Soft drink	303±16.8[e]	143.9±10.8*[e]	194.3±12.4[e]	94.5±8*[e]	117.6±10.7[e]	62.4±7*[e]
Tea + sugar	26.1±4.6	33.7±5.2[e]	19.2±3.2	19.6±6.3	8.9±2.3	8.7±2.1
Coffee + sugar	2.6±0.7	5.9±2.5	2.2±0.8	5.4±2.2	2.5±1.2	4.2±1.2
Plain milk	230.5±13.3[e]	184.5±10.2*	176.2±9.8	169.1±7.4	160.7±8.9	160.4±7.5
HF milk-based	61.2±12.6	57.2±7.5	45.1±4.8	39.6±4.2	41.1±6.2	34.3±5.2
LF milk-based	18.1±3.8[e]	13.5±3	10.1±2.2[e]	7.7±2	2.6±0.9[e]	4.1±1.3
Milk others	5.1±1.6	8.6±2.1	6.6±1.8	8.2±1.5	5.3±1.5	8.8±1.8
Vegetable juice	10.2±2.7	9.3±2.2	15.4±3.2	9.6±1.9	21.8±5.4	12.7±2.2*
Fruit juice	164.8±16.9[e]	131.4±10.7*[e]	84.7±6.2	71.7±6.2	87.6±8.8	77.3±5.4
Beer	298±25.2[e]	58.7±9.7*	229.1±16.4[e]	50.5±7.3*	181.7±17.9[e]	33.7±5.2*
Wine	18.9±5	20.1±4.8[e]	28.6±4.1	37.4±4.1	49.6±5[e]	32.8±2.8*
Alcohol others	14.1±3	27±5.3	20.2±5.3	19.2±4.2	14.5±3.1	8.3±2.4
Water	1114.5±41.6[e]	1032.2±34.3	908.2±32.4 □	1093.9±41.4*	767.1±31.9 □	886.7±27.3* □
Diet soft drink	23.4±4.4[e]	40.1±5.6	51.6±8.5	57.6±5.9	49.9±6.9	46.6±4.5
Tea	84.1±8[e]	114.4±8.8[e]	120±8.4[e]	169.8±10.3*[e]	159.8±12.5[e]	215.9±10.9*[e]
Coffee	223.2±13.9[e]	178.8±12.3[e]	437.3±16.1	355.5±15.9*	458.5±17.2	343.5±11.8*
By beverage group						
1) SSB	488.4±21.8[e]	309.7±16.9*[e]	318.6±15.9[e]	186.9±12.7*[e]	179.5±12.3[e]	122±8.3*[e]
2) Nutrient-based	489.9±25.9[e]	404.6±17.1*[e]	338.1±13.2	305.9±10.3	319.1±14.5	297.6±10.9
3) Alcohol Total	331.1±26.2[e]	105.8±14.2*	277.9±18.5	107.1±9.7*	245.8±19.2	74.8±6.3*
4) Non-caloric	1445.2±43	1365.4±34.3[e]	1517.1±33.5	1676.8±45.1*[e]	1435.4±36.5	1492.7±30.5[e]
All beverages	2754.6±60.1[e]	2185.6±44.3*	2451.8±40.8[e]	2276.9±45.5*	2179.7±40.5[e]	1987.2±32.7*[e]

Note: Results are weighted and bootstrapped.
LF, low fat; HF, high fat; SSB, sugar-sweetened beverages.
* Significantly different from estimate for men in corresponding age ($p<0.05$), using one-way ANOVA
[e] Significantly different from estimate for same sex in other age groups ($p<0.05$), using one-way ANOVA

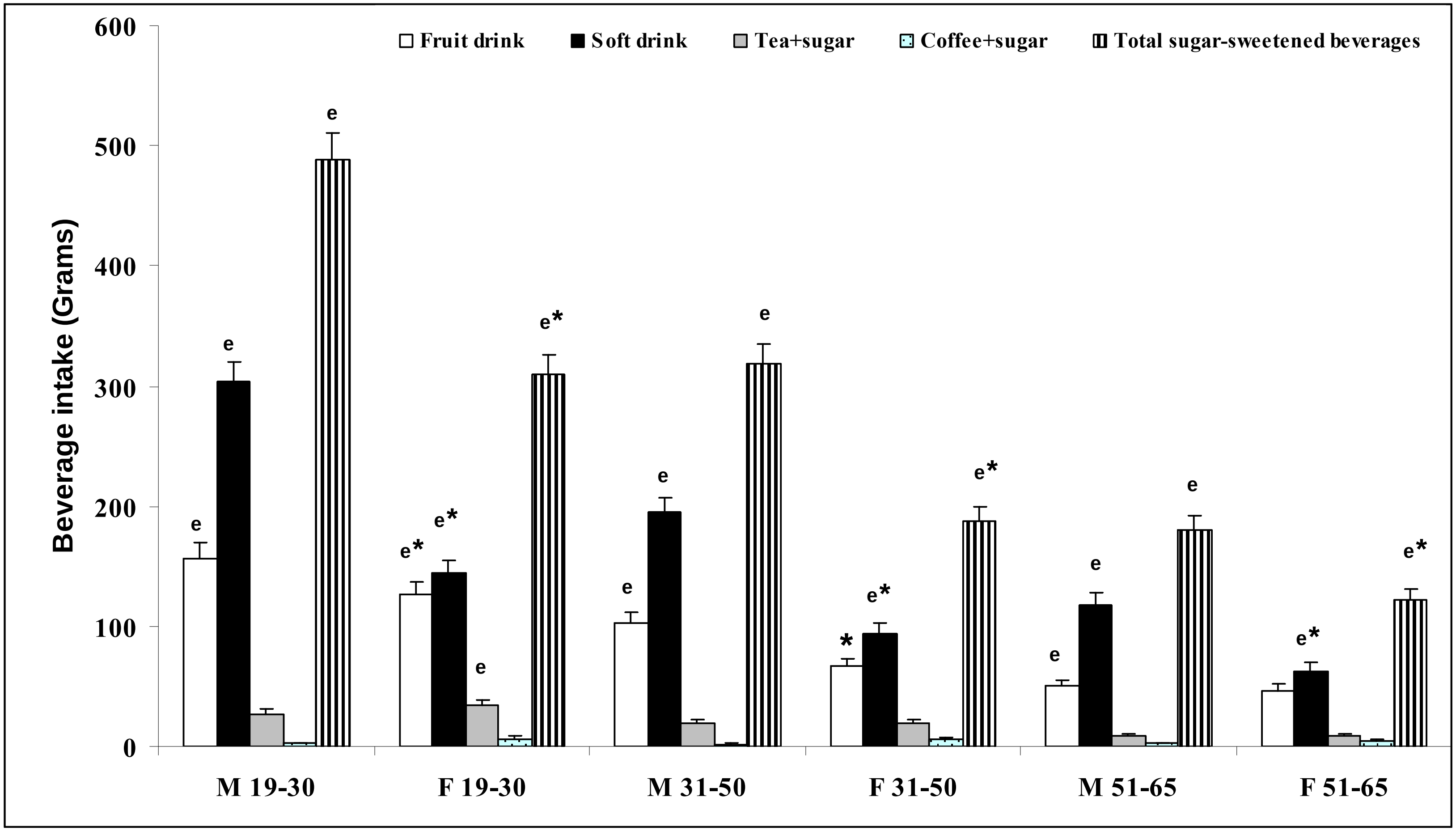


Figure 4.1. Sugar-sweetened beverage intake of Canadian adults by age/sex groups

M, males; F, females
* Significantly different from estimate for men in corresponding age ($p<0.05$), using one-way ANOVA
[e] Significantly different from estimate for same sex in other age groups ($p<0.05$), using one-way ANOVA

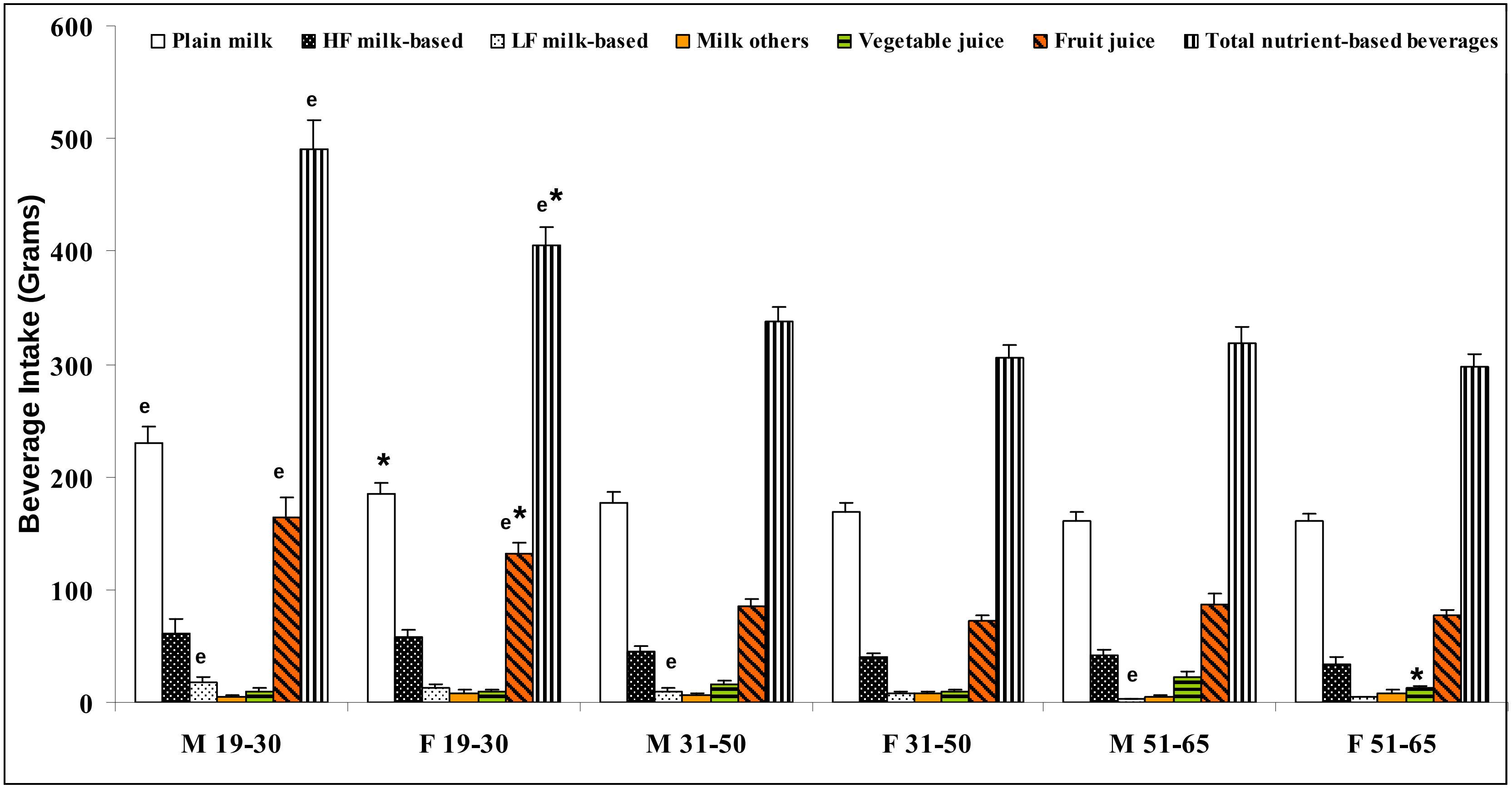


Figure 4.2. Nutrient-based beverage intake of Canadian adults by age/sex groups

LF, low fat; HF, high fat; M, males; F, females
* Significantly different from estimate for men in corresponding age ($p<0.05$), using one-way ANOVA
[e] Significantly different from estimate for same sex in other age groups ($p<0.05$), using one-way ANOVA

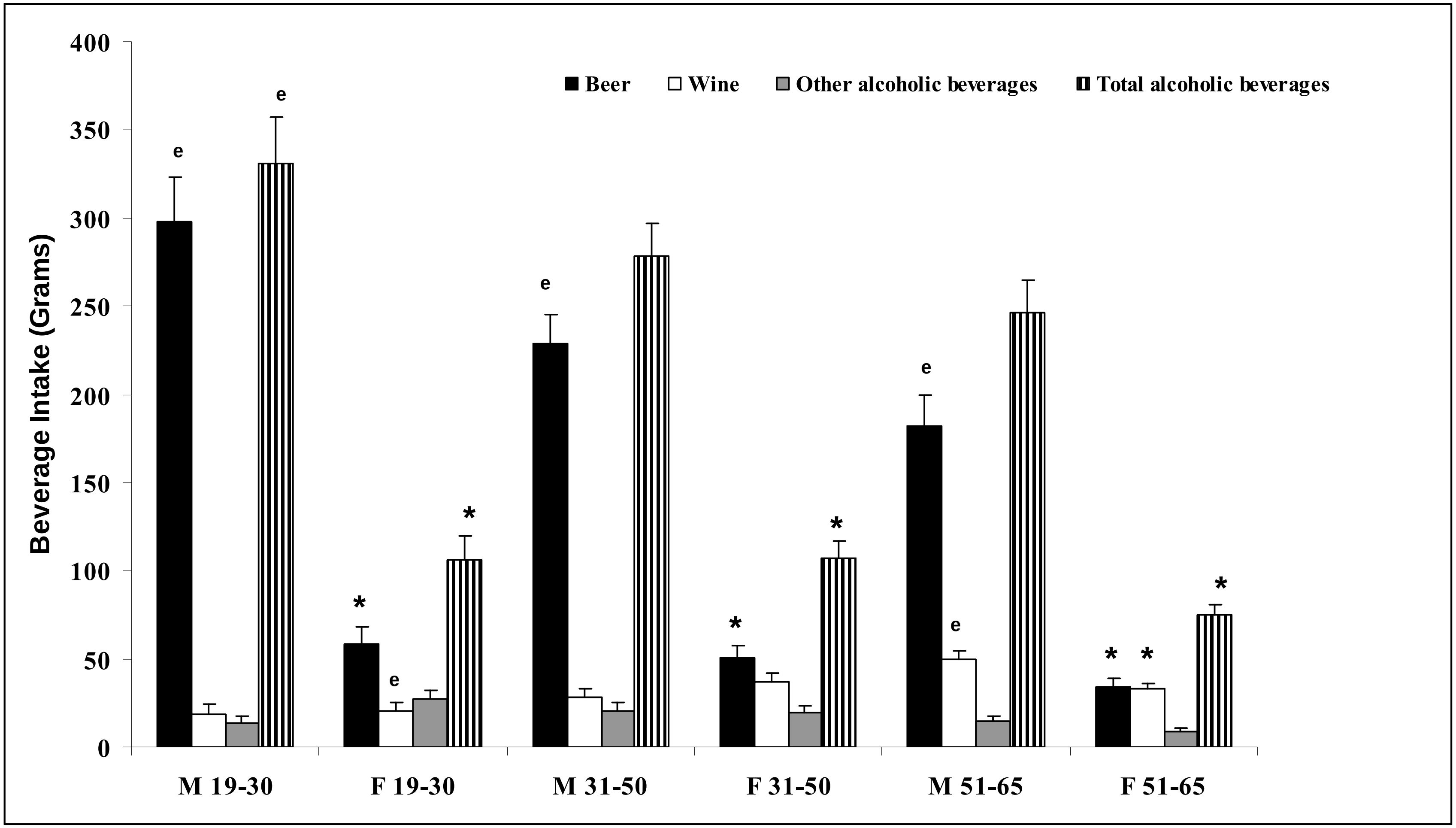


Figure 4.3. Alcoholic beverage intake of Canadian adults by age/sex groups

M, males; F, females

* Significantly different from estimate for men in corresponding age ($p<0.05$), using one-way ANOVA

[e] Significantly different from estimate for same sex in other age groups ($p<0.05$), using one-way ANOVA

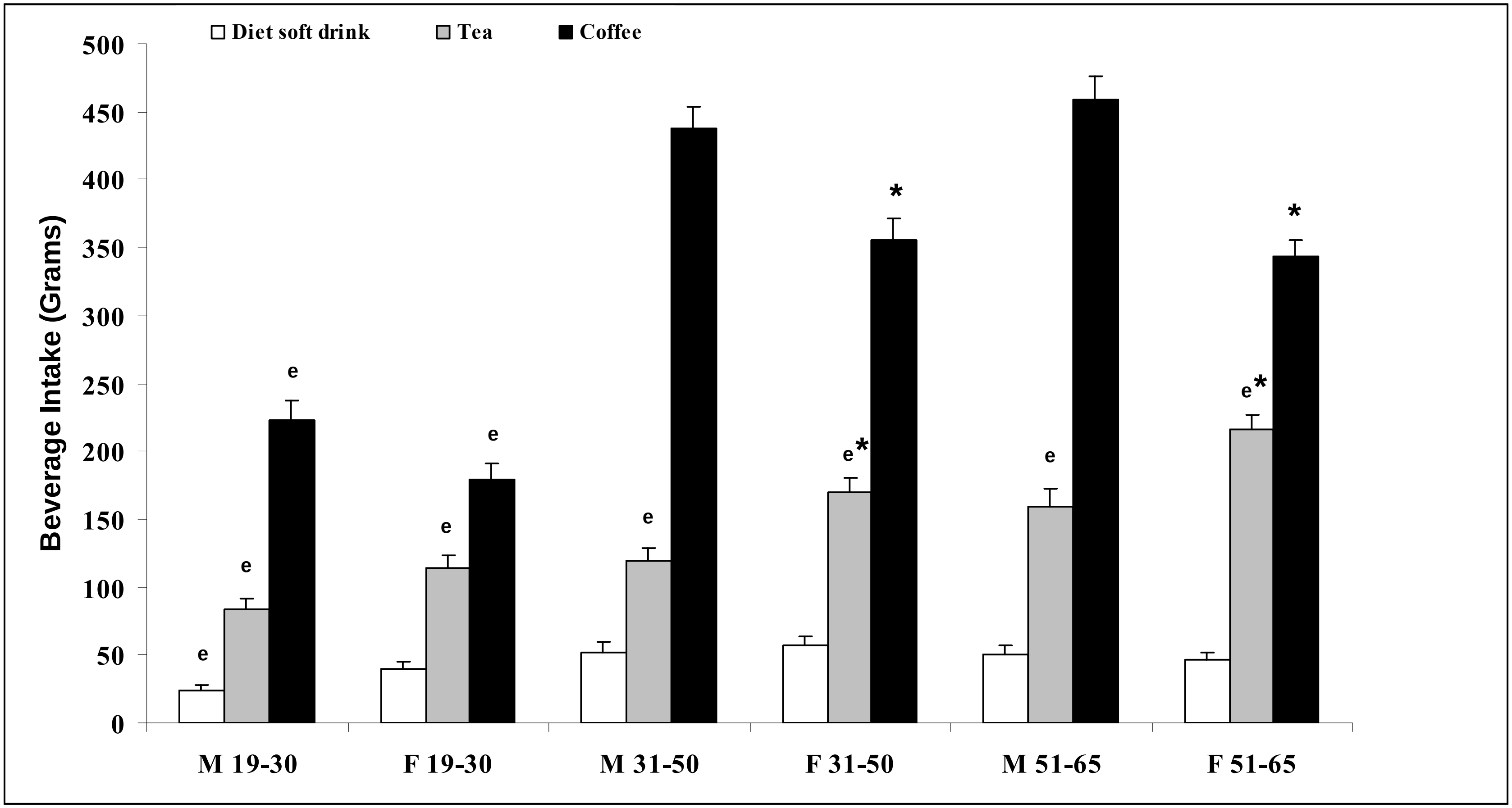


Figure 4.4. Non-caloric beverage intake of Canadian adults (excluding water) by age/sex groups

M, males; F, females
* Significantly different from estimate for men in corresponding age ($p<0.05$), using one-way ANOVA
[e] Significantly different from estimate for same sex in other age groups ($p<0.05$), using one-way ANOVA

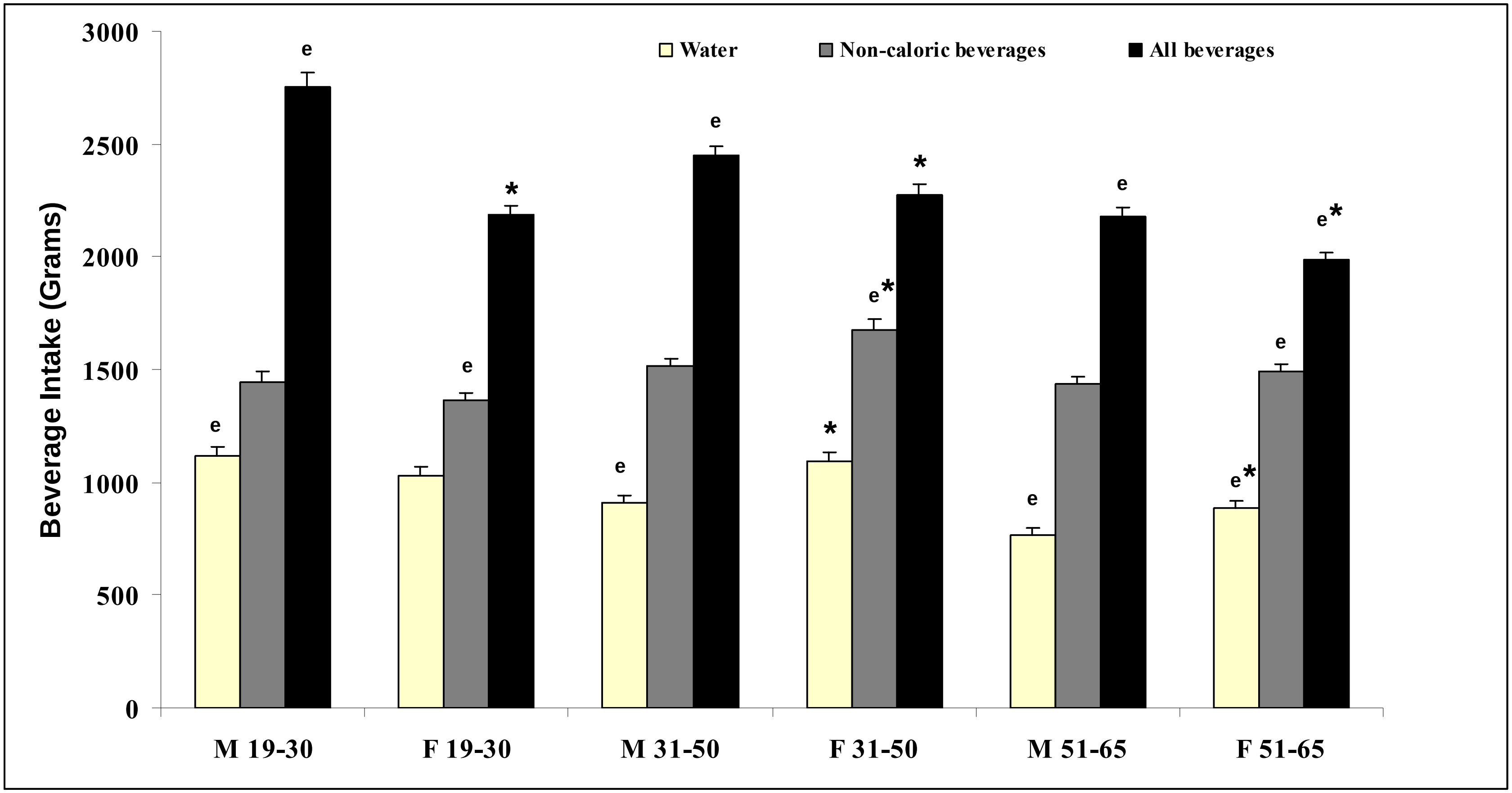


Figure 4.5. Water, total non-caloric (including water) and total beverage intake of Canadian adults by age/sex groups

M, males; F, females

* Significantly different from estimate for men in corresponding age ($p<0.05$), using one-way ANOVA

[e] Significantly different from estimate for same sex in other age groups ($p<0.05$), using one-way ANOVA

4.1.2 Contribution of Beverages to Total Energy Intake across Age/Sex Groups

The contribution of main categories of caloric beverages (sugar-sweetened, milk categories, juice, and alcoholic) to total energy intake were investigated. With the same approach explained in Section 4.1.1, the estimated energy intake from different caloric beverage groups and standard deviation at population level were calculated for 11,811 Canadian adults with plausible energy intake: 5,748 males and 6,063 females by three age groups (Tables 4.2.a, 4.2.b, and 4.2.c). Total energy from caloric beverages in men was significantly higher than in women in all three age groups ($p<0.05$). Men received significantly higher energy from sugar-sweetened beverages and alcoholic beverages than women in all age groups ($p<0.05$). The percent contribution of energy from milk beverages to total energy intake from food was significantly higher in women than men, however in terms of quantity, energy intake from milk in women was significantly lower than men particularly in younger adults.

In men, younger adults (19-30 years) had the highest total energy intake from caloric beverages (SSB, milk, juice, and alcoholic beverages) among all age groups (631 kcal/day or 21% contribution of total energy from foods and beverages). Total energy intake from caloric beverages was the lowest in the old age group (392 kcal/day or 16 % contribution in men aged 51-65 year). Moreover, alcoholic beverages had the highest contribution to total energy intake from beverages in men except for younger adults whose energy intake from SSB consumption had the highest rank. In all age groups of men, juice had the lowest contribution to energy intake among all caloric beverages.

Women aged 19-30 years had the highest intake of total energy from caloric beverages (SSB, milk, juice, and alcoholic beverages) compared to other age groups (429 kcal/day or 20% contribution of total energy from foods and beverages) and it decreased gradually in ascending age groups (261 kcal/day or 13.8 % contribution in women aged 51-65 years). In all age groups, milk had the highest contribution to energy intake from beverages, while juice and alcoholic beverages had the lowest contribution. Sugar-sweetened beverages ranked after milk in terms of their contribution to energy intake from beverages, particularly in women aged 19-30 years and 31-50 years.

The results from comparisons across age groups and in men and women are presented in Table 4.2.a, 4.2.b and 4.2.c and Figure 4.6.

Table 4.2.a. Contribution of beverages to total energy intake among Canadians 19-30 years with plausible energy intake, by sex

Sex	Male (n=1,653)		Female (n=1,590)	
	Energy (kcal)	%contribution	Energy (kcal)	%contribution
SSB	208.2 ±9.9	7.3 ±0.4	131.4 ±7.4*	6.3 ±0.3*
Milk	177.3 ±11.7	5.8 ±0.3	144.5 ±7.7*	6.9 ±0.3*
Juice	76.4 ±6.8	2.6 ±0.2	70.1 ±6.1	3.4 ±0.3*
Alcohol	169 ±13.3	5.4 ±0.4	82.9 ±12.4*	3.4 ±0.4*
Total energy from caloric beverages	630.9 ±21.1	21 ±0.6	429 ±16.6*	20 ±0.6
Total energy from food only [n]		≈79		≈80
Total energy from food and beverages	2923 ±46.5	100	2092.6 ±33.8*	100

Values are mean ± standard error of the mean (SEM), weighted bootstrapped
SSB, sugar-sweetened beverages
*Significantly different from estimate for men in corresponding age ($p<0.05$), using one-way ANOVA
[n] Estimated values calculated by subtracting energy intake of beverages from total energy intake

Table 4.2.b. Contribution of beverages to total energy intake among Canadians 31-50 years with plausible energy intake, by sex

Sex	Male (n=2,323)		Female (n=2,293)	
	Energy (kcal)	%contribution	Energy (kcal)	%contribution
SSB	141.3 ±7.1	5.1 ±0.2	81.8 ±5.4*	5 ±0.3*
Milk	133 ±6.4	4.9 ±0.2	122.9 ±5	6.2 ±0.3*
Juice	44.7 ±3.3	1.7 ±0.1	37.5 ±3.2	2 ±0.2
Alcohol	144.5 ±9.7	5.3 ±0.3	71.4 ±7.1*	3.3 ±0.3*
Total energy from caloric beverages	463.5 ±14.1	17 ±0.5	313.6 ±10.7*	15.5 ±0.4*
Total energy from food [n]		≈83		≈84.5
Total energy from food and beverages	2722.7±39.6	100	2016 ±29.2*	100

Values are mean ± standard error of the mean (SEM), weighted bootstrapped
SSB, sugar-sweetened beverages
*Significantly different from estimate for men in corresponding age ($p<0.05$), using one-way ANOVA
[n] Estimated values calculated by subtracting energy intake of beverages from total energy intake

Table 4.2.c- Contribution of beverages to total energy intake among Canadians 51-65 years with plausible energy intake, by sex

Sex	Male (n=1,772)		Female (n=2,180)	
	Energy (kcal)	% contribution	Energy (kcal)	% contribution
SSB	81.1 ±6	3.3 ±0.2	54.7 ±4.5*	2.8 ±0.2
Milk	115.9 ±6.4	4.8 ±0.2	113.7 ±5.9	6.1 ±0.3*
Juice	47.1 ±5.2	2 ±0.2	41.2 ±2.9	2.2 ±0.1
Alcohol	147.7 ±10.8	5.9 ±0.4	51.9 ±4.3*	2.7 ±0.2*
Total energy from caloric beverages	391.9 ±13.6	16 ±0.5	261.5 ±8.2*	13.8 ±0.4*
Total energy from food [n]		≈84		≈86.2
Total energy from food and beverages	2470.4 ±33.7	100	1895.3 ±22.3*	100

Values are mean ± standard error of the mean (SEM), weighted bootstrapped
SSB, sugar-sweetened beverages
*Significantly different from estimate for men in corresponding age ($p<0.05$), using one-way ANOVA
[n] Estimated values calculated by subtracting energy intake of beverages from total energy intake

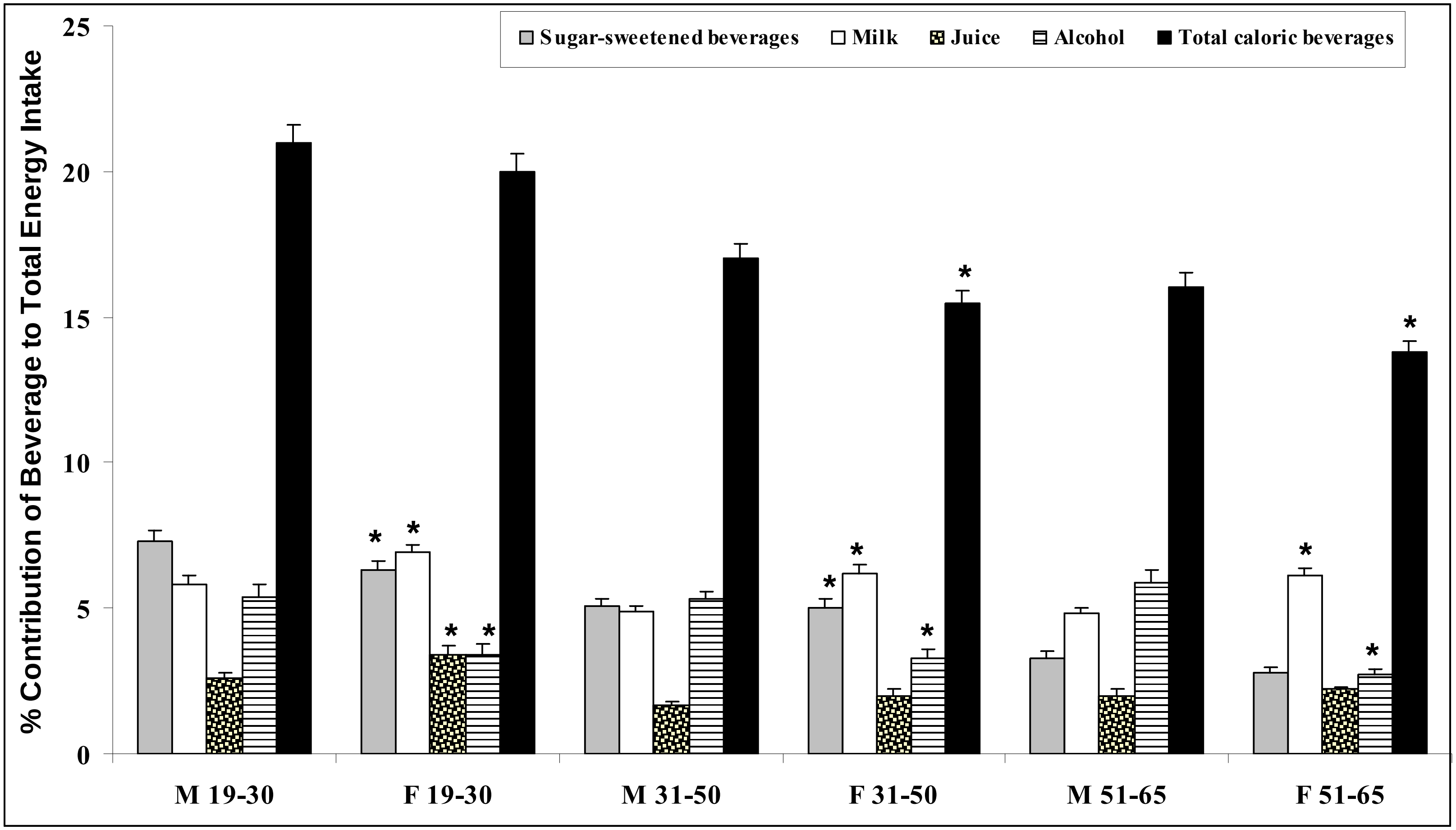


Figure 4.6. Percent contribution of caloric beverages to total energy intake of Canadian adults by sex/age groups

* Significantly different from estimate for men in corresponding age ($p<0.05$), using one-way ANOVA

4.2 Determining Patterns of Beverage Intake

In this section results of cluster analyses are presented in three sections including explorative cluster analysis, final cluster analysis, and results from statistical modeling.

4.2.1 Results of Explorative Cluster Analyses

To determine the most suitable patterns of beverage consumptions in terms of interpretability and adequacy of statistical analyses, explorative analyses were conducted in four steps.

Step 1. Cluster Analyses in Total Adult Population

Data on beverage consumption of 14,304 men and women were applied for cluster analysis (n=27 missing values). Based on the Cubic Clustering Criterion (CCC), pseudo-F-statistic,[164] and interpretability, a six cluster solution was the most appropriate pattern for Canadian adults. For sensitivity analysis, we reanalysed the data set with cluster number five and seven. However, those clusters showed neither meaningful, nor strong as six clusters. Clusters were labelled according to the dominant beverage in each cluster. The cluster with no dominant beverage (excluding water) was named “mix” which represents the general population who drink a combination of various beverages. The six beverage clusters were: coffee (n= 1,981), tea (n=6,435), water (n=867), soft drink (n=1,017), beer (n=572), and mix group (n=3,405). Consumptions of beverages (grams/day) in six clusters are summarized in Table 4.3.

Table 4.3. Intake of beverages (grams/day)* among different clusters in Canadian adults

Beverage	Cluster1 Coffee n=1,981	Cluster2 Tea[α] n=6,435	Cluster3 Mix n=3,405	Cluster4 Water n=867	Cluster5 Soft drink n=1,017	Cluster6 Beer n=572
Fruit drink	58.8±6.1	94.2±5.4	90.2±7.5	92.4±14.6	116.1±12.9	83.1±17.2
Soft drink	127.5 ±11.2	71.2±3.3	90.3±7.7	79.6±11.9	**1031.5±25.9**	222±30.7
Tea+ sugar	8±2.2	23.2±3.2	20±4.4	17.9±6.2	15.2±6.6	8.5±3.7
Coffee+ sugar	2.4±1	3.8±1.1	3.8±1.4	4.4±1.8	7.3±3	1.6±0.9
Plain Milk	171.7±10.7	185±6.4	172.4±7.8	214.2±18.2	155.4±14.2	148.5±14.7
HF milk based	45.9±7.2	42.7±3.1	47.2±7.2	53.4±9.5	56.5±12.1	32.5±6.9
LF milk based	8±2.9	9.7±1.2	6.7±1.5	5.2±1.9	18.7±4.9	12.6±9.6
Milk others	7.2±2.9	6.4±1	9.6±1.8	9.9±2.8	1.9±1.7	4.5±2.2
Vegetable juice	13.2±4.4	14.4±1.9	12.5±2.3	9.5±2.2	6.8±2.4	15.6±5.7
Fruit juice	65.5±6.8	109.6±5.7	96.9±6.9	92.9±12.1	88.7±14.3	68.9±11.7
Beer	86±8.8	49.7±3.5	80.5±7.7	142.9±23.7	57.5±8.7	**1839±53.7**
Wine	39.4±5.4	31.1±2.6	33.9±3.9	30.1±6	7.2±1.8	42.1±14.5
Other alcohol	21.9±5.9	15.3±3.1	14.9±2.4	15.4±3.2	28.6±9.3	30.5±11.3
Water	410±17.7	**462.5±8.2**	1771.5±14.8	**3908.2±83.5**	355.6±23.9	564.4±40.6
Diet soft drinks	57.1±7.7	57.5±5.1	38.9±4.5	35.7±8.4	5±2.3	29.3±14.7
Tea	62±8.1	195±7	128.2±7.9	134.1±17.3	72.3±11.7	74.1±11.7
Coffee	**1166.8±26.9**	191.7±4.6	261.5±11.1	280.3±22.5	227.3±14.8	395.1±32.3

*Values are mean (g/d) ± standard error of the mean (SEM)

[α] This cluster was named "Tea", however the predominant beverage intake in this cluster was water.

Note: the bolded-underlined numbers indicates the predominant beverage intake in each cluster.

Step 2. Cluster Analysis in Total Adult Population, Excluding Water

Similar to previous results presented in Table 4.1 and Figure 4.5, water intake is the dominant beverage intake which is also affecting the emergence of other clusters (e.g. tea and mix clusters) in this step. Therefore, it was decided to run cluster analysis again, excluding water intake (Table 4.4). Excluding water category from cluster analysis resulted in five clusters with meaningful distribution of beverages in each cluster: beer (n=712), mix (n=8,293), soft drink (n=1,051), milk (n=1,648), and coffee (n=2,573).

Table 4.4. Intake of beverages (grams/day)* among different clusters in Canadian adults, water excluded from cluster determination

Beverage	Cluster1 Beer n=712	Cluster2 Mix n=8,293	Cluster3 Soft drink n=1,051	Cluster4 Plain Milk n=1,648	Cluster5 Coffee n=2,573
Fruit drink	82.4±14.8	86.7±4.6	124.9±13.7	133.6±15	59.1±5.4
Soft drink	205±26.1	66.3±2.7	**1072.1±23.2**	90.3±8.3	107.5±9.1
Tea+ sugar	8.4±3.2	24.6±3	19.6±7.3	11.1±2.8	7.3±2
Coffee+ sugar	2.1±0.9	3.4±0.7	6.9±2.9	7.9±4.5	1.9±0.8
Plain Milk	142.6±12	93.4±2.5	120±8.9	**771.7±17.3**	137.7±6.5
HF milk based	68.6±29.4	41.5±2.7	53.6±10.7	46.6±5.8	47.4±6.5
LF milk based	17.9±9.6	7.8±1	17±4.5	11.7±2.4	6.5±2.3
Milk others	6.3±3	8.9±1	1.9±1.6	2.1±1.2	6.4±2.2
Vegetable juice	13±4.5	14.6±1.8	7.2±2.4	7.5±1.4	13.5±3.6
Fruit juice	86.7±13.4	101.2±4.5	102.8±16.4	116.8±10.8	69.1±6.3
Beer	**1723.5±45.7**	47.4±3.1	60.1±10.7	44.4±6.4	88.7±8.4
Wine	45.8±14	31.8±2.2	12.2±3.7	20.7±4.3	42.2±5.2
Other alcohol	27.6±9.2	14.7±2.4	30.2±8.6	13.3±4.4	22.2±4.9
Diet soft drinks	27.4±11.9	54.4±4.3	3.9±1.9	34.3±4.8	52.4±6.5
Tea	76.3±11.4	188.5±6.3	74.5±11.4	106.3±10.7	63.7±6.8
Coffee	381.2±27.2	176.1±4	221.9±15.8	203.4±10.3	**1108.7±21**
Total beverage	3857.6±104.5	1986.1±22	2664.9±69.3	2682.7±51.4	2663±41.8
Total % of Energy from beverages	35.9±1.1	14.1±0.2	27.8±0.7	24.9±0.5	15.3±0.4
Water[α]	942.6	1024.7±21.3	736.4±59.3	1061±42.8	828.6±30.8

*Values are mean (g/d) ± SEM and the bolded-underlined numbers indicates the predominant beverage intake in each cluster.

[α]Water was excluded in cluster analysis; however, mean intake of water was calculated after cluster analysis to assess the distribution of water intake among different clusters

Step3: Sex-specific Analysis

The idea of performing the cluster analysis in different sex groups originated from descriptive results on "beer" cluster with the smallest size. Sex distribution in this group of population with dominant beverage intake of beer was 89% and 11% in males and female, respectively. A six cluster solution for women and eight cluster solution for men was most suitable. Despite our expectation, beer cluster was emerged in both males and females in sex-specific analysis (Table 4.5.a). Moreover, two beer clusters were emerged among male population.

Table 4.5.a. Cluster analysis by sex, total adult population, water excluded from cluster determination

Sex	Cluster1 Fruit drink	Cluster 2 Soft drink	Cluster 3 Tea	Cluster 4 Coffee	Cluster 5 Mix	Cluster 6 Beer (light)	Cluster 7 Beer (heavy)	Cluster 8 Plain milk
Female n=7,463	n=411	n=447	n=1,182	n=1,212	n=4,026	n=185	-------	-------
Male n=6,814	n=482	n=658	n=699	n=850	n=2,629	n=600	n=234	n=589

Step4: Age/Sex Group Analysis

Since beer intake was higher in younger adults, it was expected to observe a better distribution by applying cluster analysis in all age/sex groups separately. Subjects were classified into four groups including males and females, aged 19-44 year and 45-65 year. However, the distribution of clusters among sex/age groups was not satisfactory (e.g. beer cluster was emerged in females with small size) (Table 4.5.b).

Table 4.5.b. Cluster analysis by sex/age, total adult population, water excluded from cluster determination

Sex/age	Cluster 1 Fruit drink	Cluster 2 Soft drink	Cluster 3 Tea	Cluster 4 Coffee	Cluster 5 Mix	Cluster 6 Beer	Cluster 7 Milk
F/19-44 years	n=288	n=323	n=407	n=514	n=2,008	n=120	-----------
M/19-44 years	n=358	n=359	n=287	n=388	n=1,578	n=371	n=315
F/45-65 years	----------	n=302	n=523	n=1,259	n=1,354	----------	n=364
M/45-65 years	----------	n=217	n=427	n=793	n=1,283	n=437	-----------

4.2.2 Final Cluster Analyses

The results of explorative analyses were reviewed. We decided to conduct cluster analyses in males and females separately (merging two beer groups in males into one group), and run the analyses in both total populations of interest and participants with plausible energy intake.

4.2.2.1 Beverage Patterns in Canadian Men

In the CCHS 2.2 sample of Canadian men using cluster analysis and excluding water as a potential beverage group, seven non-overlapping beverage patterns were determined including fruit drink, coffee, milk, beer, soft drink, tea, and a mix group with no predominant beverage consumption. The mix group had the largest sample size (n=2,629) and the lowest amount of total beverage intake (1,930 grams/day) compared to the other six clusters. The soft drink cluster had the lowest consumption (2.9 grams/day) of diet drinks compared to other clusters. The beer drinking pattern had the highest total beverage intake (3,345grams/day), as well as the highest total percent of energy intake from beverages (30.5%) in comparison with the other six groups. Percent energy intake in Canadian men characterized as tea drinkers (11.1%) was significantly lower than other beverage clusters. Meanwhile the consumption of sugar-sweetened tea (46.5grams/day) was also higher among tea (with no added sugar) drinkers (Table 4.6).

Table 4.6. Intake of beverages (grams/day)* among different clusters in Canadian male adults

Beverage	Cluster1 Fruit drink n= 482	Cluster2 Coffee n=850	Cluster3 Milk n=589	Cluster4 Beer n=907	Cluster5 Mix n=2,629	Cluster6 Soft drink n=658	Cluster7 Tea n=699	P-value
Fruit drink	**866.8±42.4***	54.1±8.7	68.4±11.6	58.9±10.3	35±3.7	78.7±12.8	36.7±6.4	<0.0001
Soft drink	143.5±19.1	146.3±16.5	131.3±17.7	183±20.3	89.5±5.4	**1117.6±30***	78.2±11.1	<0.0001
Tea+ sugar	16.3±5.9	11.1±4	17±6.9	7.4±2.5	17.8±3.1	19.1±8.9	46.5±11.4*	<0.0001
Coffee+ sugar	1.3±1	2.3±1.7	1.7±1.3	4.1±2.1	1.9±0.5	5±3.1	1.2±0.8	0.051
Plain Milk	146.4±16.8	156.2±12.1	**958±36.5***	137.2±10.3	106.6±5.5	125.8±12.2	129.2±12.1	<0.0001
High-fat milk based	67.6±15.3	44.9±11.9	39±9.6	58±21.5	47±5.7	46.3±8.3	38.4±10.4	0.163
Low-fat milk based	13.2±9.1	8.3±4.2	16.3±5	9.8±5.7	9.5±1.9	13.3±3.9	6.5±2.3	0.354
Milk others	7.3±6	1.8±1.2	0.2±0.2	5.1±2.4	6.4±1.3	2.8±2.5	15.5±7	<0.0001
Vegetable juice	13.9±5.7	19.1±9.5	10.1±3.2	9.5±3.1	19.5±4.2	8.3±3.3	15.9±4.7	0.021
Fruit juice	87.5±16	70.2±11	140.2±19.9	79.6±10.3	123.5±9.8	117.2±23.6	85.7±15	<0.0001
Beer	57±15	76.4±13.2	53.2±11.1	**1355±40.9***	40.3±4	68.3±15	35.7±7.7	<0.0001
Wine	16.3±4.8	38.2±8.9	21.2±8.6	44±9.6	37.8±4.6	11.4±4.9	14.4±4.3	<0.0001
Alcohol other	28.4±16	13.9±6.7	6.2±2	16.1±3.1	18.8±5.8	24±6.2	9.3±4.1	0.110
Diet	44.1±14.2	42.1±9.8	28.5±7.3	28.1±9.3	66.3±9.7	2.9±2.5*	18.2±4.3	<0.0001
Tea	60.5±15.7	64.6±13.8	49.7±8.9	68.6±9.2	42.7±3.5	55.8±12	**763.1±25.2***	<0.0001
Coffee	183.8±19.5	**1403±38.4***	238.9±20.1	408.6±19.2	263.5±8.9	233.5±20.8	156.6±15.7	<0.0001
Total beverage	2727.1±93.1	2882.3±70.9	2869.7±94.2	3344.6±89*	1930±32.4*	2774.4±94.3	2242.3±71*	<0.0001
% of dominant beverage to total	31%	48%	34%	40%	-------	40%	34%	--------
Total % of energy from beverages	24.1±1.1	14.4±0.7	23.8±0.7	30.5±0.8*	13.5±0.4	26.1±0.8	11.1±0.6*	<0.0001
Water$^{\alpha}$	973±68	729.7±55.8	1089.7±78	871.5±58	1003.3±32.4	844.2±79.6	791±54.6	<0.0001

*Significantly different compared to other groups (P<0.05), using one-way ANOVA

Note: Values are mean (g/d) ± SEM and the bolded numbers indicates the predominant beverage intake in each cluster.

$^{\alpha}$Water was excluded in cluster analysis; however mean intake of water was calculated after cluster analysis to assess the distribution of water intake among different clusters.

4.2.2.2 Socio-demographic and BMI Comparison across Clusters in Men

For Canadian men, total household income, income-related food security, and BMI classification (as a categorical variable) did not differ across beverage clusters. Age, ethnicity, physical activity, area of residence (urban/rural), mean BMI (as a continuous variable), smoking status, education, immigration status, marital status, mean energy intake from food, and season did differ significantly in multiple comparison among beverage clusters ($p<0.05$). However, there was no significant difference between beverage clusters in post-hoc tests of age, physical activity, area of residence (urban/rural), mean BMI (as a continuous variable), smoking status, marital status, and mean energy intake from food (Table 4.7).

Men's ages ranged from a mean of 34.6 years in the soft drink cluster to 44.5 years in the coffee drinker cluster. Larger proportion of non-white ethnicity (31.9%) and immigrants (41.9%) were in the tea drinker cluster compared to other clusters. Moreover, the coffee drinkers and soft drink cluster had lowest proportion of adults with post secondary education (49.1% and 42.8% respectively) compared to other beverage clusters. In terms of season of interview, the beer drinkers had the lowest proportion of winter participants (29.8%) compared to other clusters except for the fruit drinker cluster (31.9%).

Finally, mean BMI ranged from 26.4 in the fruit drink cluster to 27.7 in milk and tea clusters. Characteristics of patterns of beverage intake across clusters in men are presented in Table 4.7.

Table 4.7. Characteristics of each beverage pattern among Canadian male adults

Characteristics	Population		Cluster1 Fruit drink n=482	Cluster2 Coffee n=850	Cluster3 Milk n=589	Cluster4 Beer n=907	Cluster5 Mix n=2,629	Cluster6 Soft drink n=658	Cluster7 Tea n=699	P-value
Age (mean ± SEM)	40.9±0.1		35.7±0.8	44.5±0.6	37.8±0.9	39.3±0.6	42.4±0.3	34.6±0.9	44±0.8	<0.0001
Ethnicity	White	80.9 %	81.4	93.1	88.1	86.8	77	83	68.1*	<0.0001
	Non-white	19.1%	18.6	6.9	11.9	13.2	23	17	31.9*	
Total household income	Lowest	2.9%	3.7	2.4	4.8	1.3	3.1	3	2.9	0.058
	Lower middle	4.3%	3.5	5.6	2.2	4.5	3.9	4.3	6.6	
	Middle	15.8%	15.1	15.4	11.8	12.7	18	12.3	18.5	
	Upper middle	36.7%	44.6	34.1	40.8	37.2	33.8	45.1	35.8	
	Highest	40.3%	33.1	42.5	40.4	44.4	41.2	35.3	36.2	
Physical activity index	Active	19%	25.1	16.7	26.3	22.1	15.6	19.4	22.5	<0.0001
	Moderate	25.1%	30.3	21.3	25.9	25.6	26.8	20.3	21	
	Inactive	55.9%	44.6	62	47.8	52.3	57.6	60.3	56.5	
Area of residence	Urban	82.5%	83.5	80.3	78.5	77.7	85.6	76.8	85.5	<0.0001
	Rural	17.5%	16.5	19.7	21.5	22.3	14.4	23.2	14.5	
BMI (mean ± SEM) n=3,775	27.2±0.1		26.4±0.4	27.5±0.5	27.7±0.4	27±0.4	27.3±0.2	26.5±0.4	27.7±0.4	0.026
BMI categories n=3,738	Normal	34.6%	35.8	28.3	32	37.1	35.5	39.8	30.5	0.439
	Overweight	41.7%	48.9	48.6	42.2	42.2	39.6	38.2	40.9	

	Obese	23.7%	15.3	23.1	25.8	20.7	24.9	22	28.6	
Smoking status	Current smoker	31.3%	24.5	52.6	21.5	45.5	25.7	40.8	15.3	<0.0001
	Former smoker	26.4%	20.6	24.7	26.4	26.1	28.9	19.4	28.2	
	Non-smoker	42.3%	54.9	22.7	52.1	28.4	45.4	39.8	56.5	
Household food security	Secure	93.6%	92.3	92.2	97.1	94.2	93.5	92.9	93.2	0.379
	Insecure	6.4%	7.7	7.8	2.9	5.8	6.5	7.1	6.8	
Respondent education	<Secondary	15%	10.2	20.3*	13.6	16.1	13	22.9*	13.6	<0.0001
	Secondary	27.5%	29.8	30.6*	27.8	23.3	27	34.5*	24.3	
	>Secondary	57.5%	60	49.1*	58.6	60.6	60	42.8*	62.1	
Immigration status	Immigrant	23.7%	22.2	13.8	14.8	17.2	28.1	14	41.9*	<0.0001
	Non-immigrant	76.3%	77.8	86.2	85.2	82.8	71.9	86	58.1*	
Marital status	M/C	65.7%	55.2	73.3	59.9	59.3	69.5	52.3	73.7	<0.0001
	W/Sep/D	7.2%	3.8	10.7	5.1	10	7.4	5.2	4.7	
	Single	27.1%	41	16	35	30.7	23.1	42.5	21.6	
Energy intake from food (mean ± SEM)	2485.6±25.7		2823.5 ±90.1	2463.5 ±71.5	2967.4 ±83.9	2867.7 ±61.9	2164.2 ±34.8	3005.9 ±85.9	2302.3 ±77.7	<0.0001
Season	Winter	37.1%	31.9^{X}	43.4	39.7	29.8*	37.3	36.1	41.8	<0.0001
	Fall/spring	31.3%	36.5	34.1	33.4	27.2	30.4	34.3	31	
	Summer	31.6%	31.6	22.5	26.9	43	32.3	29.6	27.2	

M, married; C, common-law; W, widowed; Sep, separated; D, divorced.
*Significantly different compared to other groups (except for groups with ^{X} superscript), using one-way ANOVA for continuous variables, and using $\chi 2$ comparisons for categorical variables (P<0.05).

4.2.2.3 Beverage Patterns in Canadian Women

Using cluster analysis (excluding water as a beverage group), six distinct beverage patterns were emerged among Canadian women in the CCHS 2.2 sample. The six beverage clusters in women included fruit drink, beer, tea, coffee, soft drink, and mix group or a combination of beverages with no dominant pattern of beverage consumption.

Similar to men (Table 4.6), the mix group in women had the largest sample size (n=4,026) and lowest amount of total beverage intake (1,914 grams/day) compared with other beverage clusters (Table 4.8). The beer drinker pattern had the highest total beverage intake (3,330 grams/day) compared to other five groups. The consumption of sugar-sweetened tea (56 grams/day) was also higher among tea (with no added sugar) drinker cluster in comparison with other clusters. Unlike men, no dominant pattern of milk consumption was emerged in Canadian women, however, the consumption of plain milk (256 grams/day) was significantly higher ($p<0.05$) among fruit drink cluster compared to other five clusters. The consumption of water (624 grams/day) in soft drink cluster was significantly lower than its intake in the other clusters (Table 4.8).

Table 4.8. Intake of beverages (grams/day)* among different clusters in Canadian female adults

Beverage	Cluster1 Fruit drink n=411	Cluster2 Beer n=185	Cluster3 Tea n=1,182	Cluster4 Coffee n=1,212	Cluster5 Soft drink n=447	Cluster6 Mix n=4,026	P-value
Fruit drink	**791.1±32.9***	79±27.2	35.7±6	36.6±5.4	62±10.5	31.3±2.3	<0.0001
Soft drink	84.8±17.6	63.3±19.2	36.8±6.9	80±14.9	**905.2±37.8***	44.5±3.3	<0.0001
Tea+ sugar	20.1±7.9	3.3±2.4	56.2±19.1*	8±3.6	22.9±11.8	13.3±2	<0.0001
Coffee+ sugar	0.1±0.1	5.1±3.6	10±6.4	2.1±1.1	13.9±6.1	4.3±1.3	<0.0001
Plain Milk	255.6±33.7*	111.6±30.4	183.5±12.3	148.8±11.3	121.2±13.4	171.1±6.5	<0.0001
High-fat milk based	25.5±7.6	44.8±17.4	42.9±8.4	32.9±5.7	60.7±24.1	43.9±4.2	0.010
Low-fat milk based	2.8±1.2	5.1±3.5	5.3±2.1	9±4.5	24±10*	7.6±1.4	<0.0001
Milk others	8.8±3.4	3.4±2.5	9.6±2.5	8.8±4.3	1.9±1.7	8.9±1.5	0.176
Vegetable juice	5.2±2.3	14.6±8.1	11±2.5	13±3.9	4.5±2.4	10.5±1.8	0.087
Fruit juice	88.7±16.2	113.8±41.2	66.5±7.3	47.9±5.7	58.6±11	103.9±6.3	<0.0001
Beer	23.8±9.3	**1174.8±92.2**	18.5±4.7	32.5±6.8	15±4.1	18.7±2.8	<0.0001
Wine	16.9±6.6	55.7±40.4	25.1±5.2	34.5±6.1	12.8±4.9	35.5±3.2	<0.0001
Alcohol other	55.1±24.4	43.3±38.6	7.4±4.6	27.1±6.8	17.2±4.9	14.1±2.6	<0.0001
Diet	37.6±11	87.9±54.1	30.6±5.3	57.9±9.3	5.5±2.6	57.6±4.5	<0.0001
Tea	96.8±15.1	37.1±12.5	**751.3±18.4***	54.4±6.1	67.9±13.3	65.3±3.2	<0.0001
Coffee	164.1±22.2	389.1±66.3	124.6±10.5	**1096.2±33.8**	205.5±22.2	187.4±5.8	<0.0001
Total beverages	2716±104	**3330±175***	2367±69	2605±62	2223±89	1914±31*	<0.0001
% intake of dominant beverage to total	29%	35%	32%	42%	41%	--------	-------
Total % of energy from beverages	29.5±1.4	34.3±3.7	13.2±0.6	14.7±0.6	28.8±1.2	15.2±0.4	<0.0001
Water$^{\alpha}$	1039±92.3	1098±147.9	952±53.1	914.9±46.6	623.8±89.7*	1096.1±32.1	<0.0001

Note: Values are mean (g/d) ± SEM and the bolded numbers indicates the predominant beverage intake in each cluster.
$^{\alpha}$Water was excluded in cluster analysis; however mean intake of water was calculated after cluster analysis to assess the distribution of water intake among different clusters
*Significantly different compared to other groups (p<0.05), using one-way ANOVA.

4.2.2.4 Socio-demographic and BMI Comparison across Clusters in Women

For Canadian women, age, ethnicity, total household income, physical activity, area of residence (urban/rural), mean BMI, BMI categories, smoking status, income-related household food security, education, immigration status, marital status, and mean energy intake from food did differ significantly in multiple comparison among beverage clusters ($p<0.05$). However, there was no significant difference between beverage clusters in post-hoc tests of age, total household income, BMI categories, household food security, respondent education, and marital status (Table 4.9). Season of the interview was the only variable that did not differ significantly across six beverage clusters in multiple comparison ($p=0.856$).

Women's ages ranged from a mean of 34.8 years in the soft drink cluster to 44.6 years in the coffee and tea drinker clusters. Tea drinker cluster had a larger proportion of non-white ethnicity (27.8%) and immigrants (34.2%) compared to the other clusters. The soft drink cluster had the highest proportion of inactive participants (74.4%) compared to other clusters. As well, the proportion of former smoker women in the soft drink cluster (16.7%) was statistically lower ($p<0.05$) compared to other clusters. The proportion of female participants living in the rural area was significantly higher in the beer drinker cluster (34.2%) compared to the other five beverage clusters. The mean energy intake from food in the mix group (1,713 kcal/day) was lower than its intake in the other groups ($p<0.05$).

In women, mean BMI ranged from 26.1 in the tea drinker cluster to 28.3 in the fruit drink cluster, significantly higher that tea drinker group ($p<0.05$). Characteristics of patterns of beverage intake across clusters in women are presented in Table 4.9.

Table 4.9. Characteristics of each beverage pattern among Canadian female adults

Characteristics	Population		Cluster1 Fruit drink n=411	Cluster2 Beer n=185	Cluster3 Tea n=1,182	Cluster4 Coffee n=1,212	Cluster5 Soft drink n=447	Cluster6 Mix n=4,026	p-Value
Age (mean ± SEM)	41.8±0.1		35.7±0.8	39.1±1.4	44.6±0.6	44.6±0.4	34.8±0.9	41.7±0.2	<0.0001
Ethnicity	White	82.6%	84.2	89	72.2*	94.6	84.6	82	<0.0001
	Non-white	17.4 %	15.8	11	27.8*	5.4	15.4	18	
Total household income	Lowest	3.5%	3.3	7.5	5.4	2.4	4.4	3.1	<0.0001
	Lower middle	5.8%	7.8	3.7	6.2	6.5	10.2	5	
	Middle	19.9%	16.5	27.6	15.2	21	25.6	20.4	
	Upper middle	35%	34.5	28.3	36.6	41.6	36.7	33.1	
	Highest	35.8%	37.9	32.8	36.6	28.5	23.1	38.4	
Physical activity index	Active	17%	20	13.4	20.5	14.3	10.2*	17.4	<0.0001
	Moderate	26%	27.6	32	24.9	27	15.4*	26.2	
	Inactive	57%	52.4	54.6	53.6	58.7	74.4*	56.4	
Area of residence	Urban	82.9%	85.6	65.8*	88.6	75.8*	85.6	83.2	<0.0001
	Rural	17.1%	14.4	34.2*	11.4	24.2*	14.4	16.8	
BMI (mean ± SEM) n=4,676	26.8±0.2		28.3±1*	27.4±1.8[X]	26.1±0.4	27.1±0.3[X]	27.2±0.8[X]	26.8±0.3	0.0002
BMI categories n=4,570	Normal	46%	37.1	49	51.4	45	43.7	45.7	0.0004
	Overweight	29.6%	31	27.7	24.3	26.7	24	32.3	

	Obese	24.4%	31.9	23.3	24.3	28.3	32.3	22	
Smoking status	Current smoker	23.8%	21	50.7	12.8	42.8	41.5	19.6	<0.0001
	Former smoker	22.8%	25.2	28.6	25.1	28.7	16.7*	21.9	
	Non-smoker	53.3%	46.1	20.7	62	28.5	41.8	58.5	
Household food security	Secure	92.2%	88.4	88.9	91.6	89.2	85.8	94.1	<0.0001
	Insecure	7.8%	11.6	11.1	8.4	10.8	14.2	5.9	
Respondent education	<Secondary	14.4%	6.8	29.7	12.9	17	18.1	14	<0.0001
	Secondary	29.5%	31.2	24.9	24.2	29	35	30.6	
	>Secondary	56.1%	62	45.4	62.9	54	46.9	55.4	
Immigration status	Immigrant	22.3%	24.5	6.6	34.2*	11.9	13.3	22.8	<0.0001
	Non-immigrant	77.7%	75.5	93.4	65.8*	88.1	86.7	77.2	
Marital status	M/C	64.8%	56.7	56.1	71.5	66	55.4	65	<0.0001
	W/Sep/D	13.6%	14.1	13.5	13.2	19	11.1	12.4	
	Single	21.5%	29.1	30.3	15.2	15	33.5	22.6	
Energy intake from food (mean ± SEM)	1810±16.3		2257.4 ±80.6	2211 ±85.4	1850 ±43.6	1797 ±35.5	2231 ±71.1	1713.4 ±20.7 *	<0.0001
Season	Winter	37.2%	33	36.7	38.5	37.7	32.7	37.4	0.856
	Fall/spring	30.6%	29	31.1	30.2	31.7	37.1	30.1	
	Summer	32.2%	38	32.2	31.3	30.6	30.2	32.5	

M, married; C, common-law; W, widowed; Sep, separated; D, divorced.
*Significantly different compared to other groups (except for groups with ^{X} superscript), using one-way ANOVA for continuous variables, and using $\chi 2$ comparisons for categorical variables (P<0.05).

4.3 Results from Statistical Modeling

Stepwise logistic regression was used to evaluate the association between patterns of beverage consumption with overweight, obesity, and overweight/obesity adjusting for possible confounders including age, area of residence, season, ethnicity, marital status, immigration status, smoking habits, food security, physical activity index , respondent education level, household income, and energy intake from food. Including total energy intake as a confounder is recommended by studies,[167] since it is considered to be a potential mediator in the association between sugar-sweetened beverages and weight gain which might underestimate this association. It is believed that higher intake of liquid calories will decrease satiety and lead to more energy intake from solid food and weight gain.[167]

Twelve Analyses were conducted separately for three outcome variables (overweight, obesity, and overweight/obesity) in men and women among two samples of interest including total adult participants and participants with plausible energy intake. By applying specific weight variable created by Statistics Canada and using bootstrap method, the results are generalizable to Canadian male and female adults. However, in bootstrapping procedure, for some cases one or more parameters could not be estimated in 2-5 out of 500 Balanced Repeated Replication (BRR) replicates which is ignorable.

4.3.1 Association between Patterns of Beverage Consumption and Overweight and Obesity among Total Male Population

In this step, 6,825 men (3,775 participants with measured height and weight) were included in the logistic regression analysis. The analysis was weighted and bootstrapped to represent 10,054,227 Canadian men. Bivariate analysis showed that age, area of residence, ethnicity, physical activity, immigration status, food security, education, marital status, smoking habits, and energy intake from food were significant predictors of overweight and obesity. In the final model, no beverage clusters remained significant after adjusting for significant confounders. Results of logistic regression are summarized in Table 4.10.

Table 4.10. Adjusted odds ratio (OR) with 95% confidence interval (CI) for predictors of overweight, obesity, and overweight/obesity among Canadian men

	Overweight n=1,574	Obesity n=895	Overweight/Obesity n=2,469
	Adjusted Odds ratio(95%CI)	Adjusted Odds ratio(95%CI)	Adjusted Odds ratio(95%CI)
Age*			
Energy intake from food*[n]			
Education			
<Secondary		1.66(0.98,2.83)	
>Secondary		---------	
Physical activity			
Active		0.57(0.35,0.95)*	
Moderate active		0.50(0.31,0.81)*	
Inactive		---------	
Marital status			
Married/Common-law	1.64(1.11,2.41)*		1.54(1.09,2.17)*
Single	---------		---------
Smoking status			
Current smoker	0.49(0.32,0.75)*	0.45(0.26,0.77)*	0.46(0.31,0.71)*
Former smoker	---------	---------	---------
Non-smoker	0.69(0.45,1.05)		0.64(0.43,0.95)*
Ethnicity			
White	---------	---------	---------
Non-white	0.47(0.25,0.87)*	0.24(0.11,0.52)*	0.41(0.23,0.71)*
Household income			
Lowest		0.42(0.18,0.99)*	
Highest		---------	

Note: Age and Energy intake from food were significant predictors for outcome variable with minimal difference in odds ratio compared to the normal BMI group.
*Statistically significant, p-value<0.05
[n]Adjusting for Energy intake from food is suggested by previous studies.[167]

4.3.2 Association between Patterns of Beverage Consumption and Overweight and Obesity among Male Population with Plausible Energy Intake

In this step, 5,748 men with plausible energy intake (3,249 participants with measured height and weight) were included in the logistic regression analysis. The analysis was weighted and bootstrapped to represent 8,730,110 men. Bivariate analysis showed that age, area of residence, ethnicity, physical activity, immigration status, food security status, education, marital status, smoking habits and energy intake from food were significant predictors of overweight and obesity. In the final model, no beverage clusters appeared after adjusting for significant confounders. Results of logistic regression analyses are summarized in Table 4.11.

Table 4.11. Adjusted odds ratio (OR) with 95% confidence interval (CI) for predictors of overweight, obesity, and overweight/obesity among Canadian men with plausible energy intake

	Overweight n=1,358	Obesity n=692	Overweight/Obesit y n=2,050
	Adjusted OR (95%CI)	Adjusted OR (95%CI)	Adjusted OR (95%CI)
Age*(per one year increase)	1.02 (1.004, 1.04)*	1.03 (1.01, 1.05)*	1.02 (1.006, 1.04)*
Education			
<Secondary		1.87 (1.09, 3.22)*	
>Secondary		---------	
Physical activity			
Active		0.63 (0.36, 1.11)	
Moderate active		0.61 (0.36, 1.01)	
Inactive		---------	
Marital status			
Married/Common-law	1.65 (1.09, 2.50)*		1.54 (1.09, 2.29)*
Single	---------		---------
Smoking status			
Current smoker	0.59 (0.38, 0.93)*	0.42 (0.23, 0.75)*	0.44 (0.28, 0.68)*
Former smoker	---------	---------	---------
Non-smoker			0.66 (0.43, 1.01)
Ethnicity			
White	---------	---------	---------
Non-white	0.47 (0.24, 0.92)*	0.19 (0.08, 0.47)*	0.39 (0.21, 0.72)*
Household income			
Lowest		0.33 (0.13, 0.83)*	
Highest		---------	

*Statistically significant, p-value<0.05

4.3.3 Association between Patterns of Beverage Consumption and Overweight and Obesity among Total Female Population

In this step, 7,479 women (4,676 participants with measured height and weight) were included in the logistic regression analysis. The analysis was weighted and bootstrapped to represent 9,754,980 women. Bivariate analysis showed that age, ethnicity, physical activity, immigration status, food security status, education, marital status, smoking habits and energy intake from food were significant predictors of overweight and obesity. In the final model, "fruit drink" cluster appeared a significant predictor for overweight, obesity and overweight/obesity after adjusting for significant confounders. Results of stepwise regression are summarized in Table 4.12 and Figure 4.7.

Table 4.12. Adjusted odds ratio (OR) with 95% confidence interval (CI) for predictors of overweight, obesity, and overweight/obesity among Canadian women

	Overweight n=1,406	Obesity n=1,293	Overweight/Obesity n=2,699
	Adjusted OR (95%CI)	Adjusted OR (95%CI)	Adjusted OR (95%CI)
Beverage pattern			
Fruit Drink	1.84 (1.06, 3.19)*	2.55 (1.46, 4.47)*	2.05 (1.29, 3.25)*
Soft Drink		1.78(0.98,3.62)	
Mix	---------	---------	---------
Age*			
Energy intake from food *^{n}			
Education			
<Secondary	1.79 (1.08, 2.97)*	2.35 (1.53, 3.62)*	2.06 (1.39, 3.06)*
>Secondary	---------	---------	---------
Physical activity			
Active	0.63 (0.43, 0.94)*	0.32 (0.22, 0.47)*	0.47 (0.34, 0.65)*
Moderate active		0.65 (0.47, 0.90)*	0.79 (0.59, 1.04)
Inactive	---------	---------	---------
Non-Immigrant		1.67 (0.98, 2.84)	1.4 (0.95, 2.06)

Note: Age and Energy intake from food were significant predictors for outcome variable with minimal difference in odds ratio compared to the normal BMI group.
*Statistically significant, p-value<0.05
^{n}Adjusting for Energy intake from food is suggested by previous studies.[167]

4.3.4 Association between Patterns of Beverage Consumption and Overweight and Obesity among Female Population with Plausible Energy Intake

In the final step, 6,063 women with plausible energy intake (3,880 participants with measured height and weight) were included in the logistic regression analysis. The analysis was weighted and bootstrapped to represent 8,310,197 women. Bivariate analysis showed that age, ethnicity, physical activity, immigration status, food security status, education, marital status, smoking habits and energy intake from food were significant predictors of overweight and obesity. In the final model, "fruit drink" cluster appeared a significant predictor for obesity and overweight/obesity after adjusting for significant confounders. Results of logistic regression are summarized in Table 4.13.

Table4.13. Adjusted odds ratio (OR) with 95% confidence interval (CI) for predictors of overweight, obesity, and overweight/obesity among Canadian women with plausible energy intake

	Overweight n=1,154	Obesity n=969	Overweight/Obesity n=2,123
	Adjusted OR (95%CI)	Adjusted OR (95%CI)	Adjusted OR (95%CI)
Beverage pattern			
Fruit Drink	1.70 (0.92, 3.13)	2.24 (1.17, 4.3)*	2.06 (1.21, 3.5)*
Mix	---------	---------	---------
Age (per one year increase)	1.02 (1.02, 1.04)*	1.03 (1.02, 1.04)*	1.03 (1.02, 1.04)*
Education			
<Secondary	---------	---------	---------
Secondary	0.45 (0.25, 0.81)*	0.44 (0.26, 0.75)*	0.43 (0.27, 0.69)*
>Secondary	0.44 (0.26, 0.74)*	0.36 (0.22, 0.61)*	0.39 (0.26, 0.60)*
Physical activity			
Active	0.68 (0.44, 1.05)	0.28 (0.18, 0.44)*	0.48 (0.33, 0.70)*
Moderate active		0.59 (0.41, 0.87)*	0.76 (0.56, 1.04)
Inactive	---------	---------	---------
Food secure	1.76 (0.99, 3.13)		
Non-white		0.55 (0.33, 0.93)*	

*Statistically significant, p-value<0.05

Overall, women with dominant pattern of fruit drinks had significantly higher odds for overweight (OR=1.84, 95% CI: 1.06-3.20, P=0.031), obesity (OR=2.55, 95% C.I: 1.46-4.47, p=0.001), and overweight/obesity (OR=2.05, 95% CI: 1.29-3.25, p=0.002), compared with women with no dominant beverage intake. However, women with plausible energy intake whose dominant beverage pattern was fruit drinks didn't have higher odds for overweight (OR=1.70, 95% CI: 0.92-3.13, P=0.087), yet they had higher odds for obesity (OR=2.24, 95% C.I: 1.17-4.31, p=0.016), and overweight/obesity (OR=2.06, 95% CI: 1.21-3.50, p=0.007), compared with women with no dominant beverage intake (Figure 4.7).

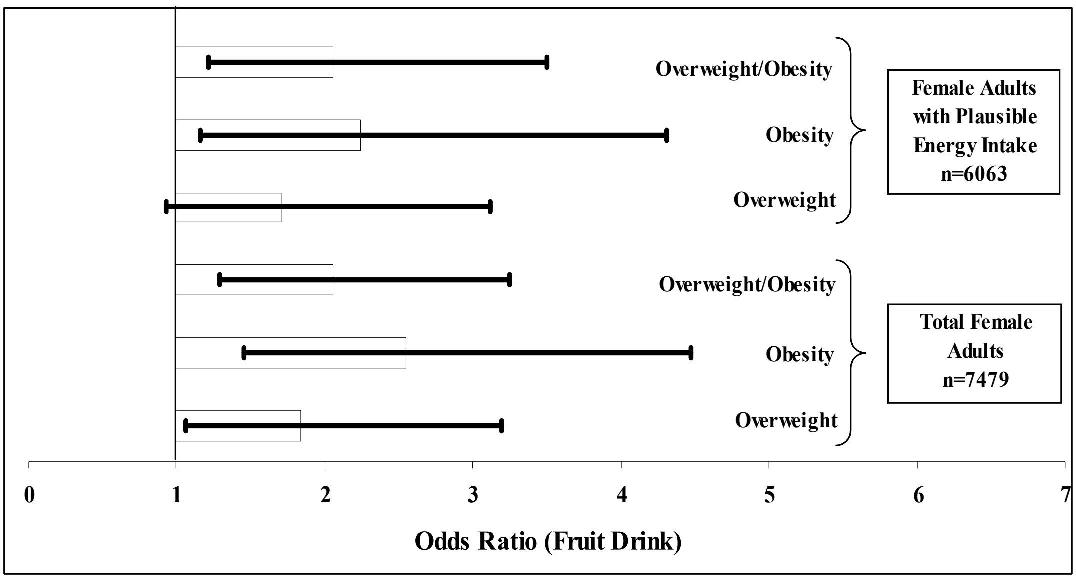


Figure 4.7. A summary of adjusted odds-ratio for dominant pattern of fruit drink (compared with moderate intake of all beverages) as a predictor for overweight and/or obesity in Canadian women

Chapter5

DISCUSSION

In this Chapter, descriptive data on beverage intake and BMI of Canadian men and women is discussed and compared with the equivalent data on American adults. Further, results from cluster analysis on patterns of beverage consumption of Canadian men and women are discussed in this Chapter. In addition, demographics and BMI comparisons across clusters are interpreted. The final results of logistic regression, interpretations, and evidence from other studies are presented. Finally, limitation of the study and recommendations for further research are explained.

5.1 Descriptive Data

In this section, we interpreted our data with available published data in Canada and the United States. The initial descriptive results of our study are similar to the descriptive report by Garriguet[11] who used CCHS 2.2 data. However, to address the objectives of our study we had a more comprehensive approach. Beverages were categorized in 17 groups and adults older than 65 years were excluded in our analysis.

To our knowledge, the study by Storey et al.[7] is the only study that reports average beverage consumption in US across different age and sex groups. This study and other reports on prevalence of overweight and obesity are based on data from the National Health and Nutrition Examination Survey (NHANES). The NHANES program includes a series of complex, stratified, multistage clustered, cross-sectional, nationally representative of the US population surveys that began in 1960.[168] Beginning in 1999, the NHANES data are released in 2-year cycles without a break, examining samples of about 5,000 persons each year.[168] To compare beverage consumption and BMI status of Canadian and American adults, we used studies based on the comparable NHANES data released in 1999-2002,[7] and 2003-2004.[169]

5.1.1 Beverage Consumption of Canadian Men Compared to That of American Men

Results from Duffey and Popkin's study[5] on beverage consumption of the US population showed that water was the most frequently consumed beverage which was similar to Canadians. After water, the most consumed beverages by Americans were coffee, soda, whole-fat milk, fruit juice, and alcoholic beverages, respectively.[5]

Similar to our results, Storey et al.[7] using data from the NHANES reported that soft drinks were the most consumed beverage by young adult men aged 20-39 years; however, the amount of consumed soft drinks was more than twice the intake of their Canadian counterparts. After soft drinks the top three beverages consumed by young adult men were coffee, tea, and milk. The younger American men drank more tea than corresponding Canadians; however they had less consumption of fruit juice, fruit drink, and milk.[7]

Parallel with our data on Canadian middle aged adults, coffee consumption replaced soft drink among American men aged 40-59 years. In American men aged 60 years and over, milk was the second most consumed beverage (after coffee), followed by tea and soft drinks (almost the same amount as Canadian middle age adults). The average consumption of milk, fruit juice, and fruit drink remained relatively steady in American men across age groups. This is different from the observed negative trend in Canadian adults across ascending age groups. Similar to Canadians, energy intake from beverage sources was the highest in younger American adults aged 20-39 years compared to other age groups. It declined by ascending age groups.[7]

There were some limitations in comparing our data with data from the study by Storey et al..[7] In the US study, there was no report on water and alcoholic beverages consumption, however energy from wine and beer were included in the non-beverage category to account for all sources of energy.[7] Moreover, there was no report of energy intake from different beverages to compare to those of Canadian adults. Finally, the age groups in the study by Storey et al. were different from our study.

5.1.2 BMI Comparison among Canadian and American Men

According to the data from CCHS 2.2, mean BMI of Canadian men aged 18 years and over in 2004 was 27.2. The prevalence of overweight among men was 42% and 22.9% of

Canadian men were obese.[127] It means that only 35% of men were in the normal or underweight categories. The prevalence of overweight/obesity was the highest among older adults (aged 55-74 year), more than twice of the rate in younger adults aged 18-24 years.[127]

Data from the NHANES in 2004 revealed that 39.7 % of American men were overweight and 31.1% of men were obese.[169] The prevalence of overweight was higher among Canadian men compared to Americans, while obesity in US had higher prevalence than in Canada. Similar to Canadians, younger adults (aged 20-39 years) had the lowest prevalence of overweight/obesity, while it was the highest among middle age adult men (aged 40-59 years).[169]

5.1.3 Beverage Consumption of Canadian Women Compared to That of American Women

Results from Storey et al.[7] showed that soft drinks were the most consumed beverage by American women aged 20-39 years which was different from Canadian young women where the most consumed beverages after water were plain milk and coffee. After soft drinks the top two beverages consumed by younger adult American women were coffee and milk beverages.[7]

Among American women aged 40-59 years, coffee consumption exceeded soft drink; the highest amount across all age groups of women. This pattern was similar to Canadian middle aged women. Among American women aged 60 years and over, milk beverages were the third consumed beverage after coffee and tea which was similar to that of older Canadian women. Comparable to Canadians, average consumption of milk remained relatively steady in American women across the age groups.

Similar to Canadian counterparts, energy intake from beverage sources was the highest in American women aged 20-39 years. This declined across age groups in women from younger to older adults.[7]

5.1.4 BMI Comparison among Canadian and American Women

Based on data from CCHS 2.2, mean BMI of Canadian women aged 18 years and over in 2004 was 26.7; approximately 30% of women were categorized as overweight and 23.2% of Canadian women were obese.[127] The prevalence of overweight/obesity was the highest among older adult women (aged 45 years and over); less than twice of the rate in younger adult Canadian women (aged 18-24).[127]

Data from the NHANES in 2004 showed that 28.6 % of American women aged 19 years and older were overweight and 33.2% of women were obese.[169] Similar to men, the prevalence of overweight was higher among Canadian women compared to Americans, while obesity in US female adult population had higher prevalence than in Canadian women. Like in Canada, younger American adult women (aged 20-39 years) had the lowest prevalence of overweight/obesity (51.7%), while the rate was higher (more than 68%) among older adult women aged 40 years and over.[168]

5.2 Cluster Analysis

Explorative cluster analyses allowed determining the most suitable clusters representing beverage consumption patterns of Canadian adults. Seven distinct beverage clusters and six clusters emerged in Canadian men and women respectively. To our knowledge, there is no equivalent study who characterized beverage intake patterns using cluster analysis in the US adult population or any other developed country to allow us comparison. All studies based on the NHANES used traditional method of examining mean intakes across age groups.

Our study showed that the mean consumption of the dominant beverage in each cluster was higher in men than in women. This finding was predictable, hence, descriptive data on food and beverage intakes of Canadians indicates men usually eat and drink more than women.[11, 170] Sex differences in body size and nutritional needs might explain higher intake of dominant beverage in men compared to women.

The emergence of beer cluster in both men and women where the intake of all other beverages were higher than the other clusters indicates popularity of this beverage. There is no data in Canada to examine the trends in beer consumption. However, Statistics Canada food disappearance data indicates decrease in availability of beer (adjusted for losses) in the last 28 years. The per capita availability of beer for Canadians aged 15 and over decreased from 99.69 liters/year in 1981 to 77.30 liters/year in 2009.[171] Data from the NHANES indicates that beer is one of the top ten sources of energy intake among Americans.[113] Whether alcoholic beverages, specifically beer, has an impact on weight gain is still controversial and needs more investigation.[110,113,114]

There was no milk cluster among women; however the milk intake among those who drank fruit drinks was significantly higher than other clusters.

Among all clusters, the mix group had the lowest consumption of total beverages in both men and women. In terms of proportion of energy intake from beverages, there was no significant difference among beverage clusters in women. In men though, tea drinkers had the lowest proportion of energy intake from beverages, and beer drinkers had the highest proportion of energy intake from beverages.

5.3 Relationship between Beverage Intake Patterns and Overweight/Obesity

Across seven beverage clusters in men of our study, there was no significant difference between mean BMI and BMI categories of participants in each cluster. Among six beverage clusters in women, the mean BMI of fruit drink consumers was significantly higher than mean BMI in tea drinker and mix groups. To evaluate the relationship between beverage patterns and overweight/obesity, results from logistic regression showed that none of the beverage clusters in men is a significant predictor for overweight, obesity, or overweight/obesity. In women, adjusting for total energy intake and other possible cofounders, the fruit drink cluster was a significant predictor for overweight, obesity, and overweight/obesity compared to mix group. However, it was not a significant predictor for overweight in women after excluding participants with implausible energy intake.

In addition to the worldwide studies evaluating prevalence of overweight/obesity and related risks of adiposity, some studies have evaluated the BMI of Canadian adults based on measured height and weight from CCHS 2.2 data,[1,127,134] besides the studies based on self-reported height and weight. Moreover, initial descriptive analysis of beverage intake of Canadian adults has been made among different age/sex groups, based on CCHS 2.2.[11] However, to our knowledge, there is no study exploring patterns of beverage intake of Canadian adults and investigating the association of sugar-sweetened beverages and BMI at national scale.

Out of 15 studies on the association of SSB and BMI (Table 2.2), only four studies evaluated male and female population individually,[77,78,80,81] one study on men only,[84] and two studies investigated this relationship in women.[79,83] Among those 15 studies on adults (Table 2.2), only Newby et al.[82] applied cluster analysis to evaluate dietary pattern, however they didn't split male and female, and beverage consumption was not evaluated separately. Duffey and Popkin[5,156] used cluster analysis to evaluate patterns and trends of beverage consumption of American adults, yet no investigation was made about its relationship with BMI. A recent review of studies evaluating the relationship between sugar sweetened beverages and risk of obesity, metabolic syndrome and type 2 diabetes revealed that sugar sweetened beverages including fruit drinks are associated with weight gain.[167] Further, association with metabolic syndrome and type 2 diabetes indicates the health risks related to high consumptions of sugar sweetened beverages.[167]

5.4 Potential Mechanisms for a Fruit Drink Association with Overweight and/or Obesity in Women

Multiple studies suggested that fruit drinks which are sugar sweetened beverages, may contribute to overweight and obesity because of excess calorie intake from sugar solutions which offer low satiety.[67-73] However, short-term human studies[172-174] and studies in experimental animals[175] show that this extra energy provided by sugar-sweetened beverages does not affect subsequent food and energy intake from solid foods.[83] Liebman et al.[78] believe that "metabolically, men may be better able to maintain caloric balance even with the higher energy intakes associated with the consumption of these larger portion sizes." This might explain the positive association of fruit drink pattern and overweight/obesity in women but not in men.

Further, total energy intake partly mediates the effect of sugar-sweetened beverages on weight.[167] In terms of energy intake from food in our study, Canadian women with dominant patterns of mix beverage consumption had significantly lower energy intake than other groups. This can explain the positive association between fruit drink cluster and overweight/obesity, as the mix group was the reference in logistic regression analysis.

Evidence from some cross-sectional studies[77-80, 167] shows positive relationship between sugar-sweetened beverages and BMI, body weight, and obesity; however, they didn't use cluster analysis to assess beverage patterns. Moreover, the impact of under reporting the energy intake was considered in our study by excluding participants with implausible energy intake. As well, our study had larger number of participants compared to other four studies. Weight and height were measured in our study while data of other studies were based on self-reported or measured weight and height.

5.5 Limitations

Our study like other studies based on large survey data may have some limitations due to the study design in survey data and analytical approaches. There are a number of limitations in using CCHS 2.2 data that affects the results of this research. Cross-sectional data does not allow us to determine the temporal relationship between beverage intake and adults' weight. In general, using survey data may restrict the investigation on other factors that may impact the relationship. However, in our statistical models, we were able to include all possible confounders that might affect the relationship between patterns of beverage intake and BMI. Total energy intake is one of the main factors that partly mediates the association between sugar-sweetened beverages consumption and BMI. Therefore, adjusting for energy intake may seriously affects the results. The positive association between obesity, overweight/obesity and dominant pattern of fruit drink intake after adjusting for total energy intake and excluding implausible energy intake supports the notion that this association is not mediated through energy intake.[175]

Data from a single 24-hour recall does not reflect day to day variation in one's usual pattern of dietary intake; however, when we deal with mean intakes, the mean intake from one-day 24-hour data is similar to the mean intake from statistically adjusted mean intake "usual intake" from two-day 24-hour recall.[19]

At the end, there is no gold standard for determining the number of clusters in cluster analysis (K means method).[144] The subjective decision needs to be made based on the existing evidence. We determined the best cluster solution by plotting Pseudo-F statistics and CCC, and identifying meaningful emerging beverage consumption patterns that is interpretable both statistically and in the light of most recent literature. In addition, the large sample size (n=14,304) allowed us to identify discrete beverage patterns, and it was less likely to commit type II error by narrowing confidence intervals. Hence, such an approach is more comprehensive in identifying patterns of beverage consumption in population than regular analysis when only mean intake of beverages is determined.

Chapter 6

CONCLUSIONS

Evidence from other studies indicates considerable increase in consumption of sugar-sweetened beverages globally. This is the first study that reports Canadian women with dominant pattern of sugar-sweetened beverages, mainly from fruit drinks, are more at risk of overweight and/or obesity compared to those who have a relatively moderate intake of all beverages. This finding was in consistent with our study hypothesis. Although the results of this study are based on cross-sectional data, the presence of positive association between fruit drink and overweight/obesity after controlling for potential mediators in a nationally representative sample indicates the importance of our findings. Longitudinal studies with adequate follow-up periods are needed to elucidate the casual relation and initiation of disease such as metabolic syndrome and type 2 diabetes by high consumption of soft drinks. However, evidence from current literature justifies the need for intervention strategies and policy initiations to target predisposing factors for overweight and obesity including high consumption of sweetened beverages.

Fruit drinks contain similar amount of calories as carbonated soft drinks. However, most preventive strategies and health messages have targeted limiting the intake of carbonated soft drinks mainly in children. Limiting the availability of sweetened beverages in food market, and replacing by health alternative beverages,[167] are examples of suggested possible policy initiations in recent literature. A critical appraisal of literature is needed to identify effective strategies in order to increase public awareness on the health risks related to high consumption of sweetened beverages and limit its availability and consumption.

REFERENCES

1. Shields M, Tjepkema M. Trends in adult obesity. Health Reports (statistics Canada, Catalogue 82-003) 2006; 17(3): 53-60.
2. Garrow JS. Obesity and related disease. London: Churchill Livingstone; 1998. p. 1-16.
3. Gilmore J. Body mass index and health. Health Reports (statistics Canada, Catalogue 82-003) 1999; 11(1): 31-34.
4. World Health Organization. Obesity and overweight. Fact sheet No. 311. September 2006. Available at URL: http://www.who.int/mediacentre/factsheets/fs311/en/
5. Duffey KJ, Popkin BM. Adults with healthier dietary patterns have healthier beverage patterns. *J Nutr*. 2006; 136:2901-7.
6. Bleich SN, Wang YC, Wang Y, Gortmaker SL. Increasing consumption of sugar-sweetened beverages among US adults: 1988–1994 to 1999–2004. Am J Clin Nutr 2009;89:372–81.
7. Storey ML, Forshee RA, Anderson PA. Beverage consumption in the US population. J Am Diet Assoc. 2006; 106:1992-2000.
8. Malik VS, Schulze MB, Hu FB. Intake of sugar-sweetened beverages and weight gain: a systematic review. Am J Clin Nutr 2006;84:274-88.
9. Brownell KD & Frieden TR. Ounces of prevention—The public policy case for taxes on sugared beverages. The New England Journal of Medicine. 2009; 360:18:1805−1808.
10. Allison, DB & Mattes, RD. Nutritively sweetened beverage consumption and obesity: The need for solid evidence on a fluid issue. JAMA, 2009; 301:3: 318−320.
11. Garriguet D. Beverage consumption of Canadian adults. Health Reports. 2008;19(4):22-29.
12. Langseth L. Nutritional epidemiology; Possibilities and limitations. ILSI Europe Concise Monograph Series; 1996. p. 1, 2.
13. Ahrens W, Pigeot I. Handbook of epidemiology. Mackerras D, Margetts BM. Nutritional Epidemiology. Germany: Springer; 2005. p. 1000.
14. Willett W. Nutritional Epidemiology. New York: Oxford University Press; 1998. p. 4.

15. Gibney MJ, Margetts BM, Kearney JM, Arab L. Public health Nutrition. Nelson M, Beresford SAA, Kearny JM. Nutritional epidemiology. Oxford: Blackwell Science; 2004. p. 26-29.
16. Gibson RS. Principles of nutritional assessment. New York: Oxford University Press; 2005. p. 4-7. p. 105-122, and p. 259-274.
17. Rhodes DG, Moshfegh A, Cleveland L, Murayi T, Baer D, Sebastian R, & Perloff B. (2004). Accuracy of 24 hour dietary recalls: preliminary results from USDA AMPM Validation Study. Federation of the American Society of Experimental Biology (FASEB) Journal; 2004: 18, A111.
18. Blanton CA, Moshfegh AJ, Baer DJ, Kretsch MJ. The USDA Automated Multiple-Pass Method Accurately Estimates Group Total Energy and Nutrient Intake. American Society for Nutrition J. Nutr. 2006; 136:2594-2599.
19. Canadian Community Health Survey, Cycle 2.2, Nutrition. A Guide to Accessing and Interpreting the Data. 2004, Cat.: H164-20/2006E-PDF. ISBN: 0-662-43554-0, page 10 and page 32-33. Available at URL: http://www.hc-sc.gc.ca/fn-an/surveill/nutrition/commun/cchs_guide_escc-eng.php
20. Briefel RR, Sempos CT, McDowell MA, Chien S, Alaimo K. Dietary methods research in the third National Health and Nutrition Examination Survey: underreporting of energy intake.Am J Clin Nutr 1997;65(4 Suppl):1203S-1209S.
21. Johansson G, Wikman A, Ahrén AM, Hallmans G, Johansson I. Underreporting of energy intake in repeated 24-hour recalls related to gender, age, weight status, day of interview, educational level, reported food intake, smoking habits and area of living. Public Health Nutr 2001;4(4):919-27.
22. Livingstone MB, Black AE. Markers of the validity of reported energy intake. J Nutr 2003;133(Suppl):895S-920S.
23. Centers for Disease Control and Prevention. Definition of Overweight and Obesity. Available at URL: http://www.cdc.gov/nccdphp/dnpa/obesity/defining.htm
24. Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. N Engl J Med. 2003;348(17):1625-38.

25. World Health Organization. Obesity: Preventing and managing the global epidemic. Report of a WHO consultation. WHO Technical Report Series 894(3), i-253. Geneva: WHO; 2000.
26. Kopelman P. Health risks associated with overweight and obesity. Short science review. Foresight tackling obesities: Future choices. Obesity Reviews. 2007; 8(s1): 13-17.
27. Isomaa B, Almgren P, Tuomi T, Forsen B, et al. Cardiovascular morbidity and mortality associated with the metabolic syndrome. Diabetes Care. 2001; 24: 683-689.
28. American Obesity Association. AOA Fact sheets: Health effects of obesity. Washington DC: AOA; 2005.
29. Ruotolo G, Howard BV, Robbins DC. Chapter 10: Dyslipidemia of obesity. In: Caro JF, Obesity: Endotext. com; 2003.
30. Gibson, GJ. Obesity, respiratory function and breathlessness. Thorax. 2000; 55(Suppl 1): S41-S44.
31. Loube DI, Loube AA, Erman MK. Continuous positive airway pressure treatment results in weight loss in obese and overweight patients with obstructive sleep apnea. Journal of the American Dietetic Association. 1997; 97: 896-897.
32. Rimm EB, Stampfer MJ, Giovannucci E, Ascherio A, et al. Body size and fat distribution as predictors of coronary heart disease among middle-aged and older US men. American Journal of Epidemiology. 1995; 141(12): 1117-1127.
33. Jood K, Jern C, Wilhelmsen L, Rosengren A. Body mass index in mid-life is associated with a first stroke in men: A prospective population study over 28 years. Stroke. 2004; 35: 2764.
34. Gortmaker SL, Must A, Perrin JM, Sobol AM, Dietz WH. Social and economic consequences of overweight in adolescence and young adulthood. The New England Journal of Medicine. 1993; 329: 1008-1012.
35. Balen A, Anderson RA. Impact of obesity on female reproductive health: British Fertility Society, policy and practice guidelines. Human Fertility. 2007; 10(4): 195-206.
36. Wagner DR, Heyward VH. Measures of body composition in blacks and whites: a comparative review. Am J Clin Nutr. 2000;71(6):1392-402.

37. Division of nutrition, Physical activity and obesity, National Center for Chronic Disease Prevention and Health Promotion. Available at URL: http://www.cdc.gov/nccdphp/dnpa/bmi/adult_BMI/about_adult_BMI.htm
38. Sweeting HN. Measurement and definitions of obesity in childhood and adolescence: a field guide for the uninitiated.Nutr J. 2007, 26;6:32.
39. Health Canada. Canadian Guidelines for Body Weight Classification in Adults. Health Canada, Ottawa. 2003. Available at URL: http://www.hc-sc.gc.ca/fn-an/nutrition/weights-poids/guide-ld-adult/cg_quick_ref-ldc_rapide_ref_e.html
40. Wang YC, Bleich SN, Gortmaker SL Increasing caloric contribution from sugar-sweetened beverages and 100% fruit juices among US children and adolescents, 1988-2004. Pediatrics. 2008 Jun;121(6):e1604-14.
41. Sun SZ, Empie MW. Lack of findings for the association between obesity risk and usual sugar-sweetened beverage consumption in adults –A primary analysis of databases of CSFII-1989–1991, CSFII-1994–1998, NHANES III, and combined NHANES 1999–2002. 2007, Food and Chemical Toxicology 45 (2007) 1523–1536.
42. Block G. Food contributing to energy intake in the U.S.: data from NHANES III and NHANES 1999–2000. J Food Comp Anal 2004;17: 439–47.
43. Nielsen SJ, Popkin BM. Changes in beverage intake between 1977 and 2001. Am J Prev Med 2004;27:205–10.
44. Ludwig DS. The glycemic index: physiological mechanisms relating to obesity, diabetes, and cardiovascular disease. JAMA. 2002;287:2414-23.
45. Tahmassebi JF, Duggal MS, Malik-Kotru G, Curzon ME. Soft drinks and dental health: a review of the current literature. J Dent. 2006;34:2-11.
46. Rampersaud GC, Bailey LB, Kauwell GPA. National survey beverage consumption data for children and adolescents indicate the need to encourage a shift toward more nutritive beverages. JAmDiet Assoc 2003; 103:97–100.
47. Almiron-Roig E, Drewnowski A. Hunger, thirst, and energy intakes following the consumption of caloric beverages. Physiol Behav 2003; 79:767–73. and Rolls BJ, Barnett RA. Volumetrics. A systematic lifetime approach to eating. New York, NY: Harper Collins, 2000.

48. van Staveren WA, Steijns JM, de Groot LC. Dairy products as essential contributors of (micro-) nutrients in reference food patterns: an outline for elderly people. J Am Coll Nutr. 2008 Dec;27(6):747S-54S.
49. Elwood PC, Givens DI, Beswick AD, Fehily AM, Pickering JE, Gallacher J. The survival advantage of milk and dairy consumption: an overview of evidence from cohort studies of vascular diseases, diabetes and cancer. J Am Coll Nutr. 2008 Dec;27(6):723S-34S.
50. Huncharek M, Muscat J, Kupelnick B. Colorectal cancer risk and dietary intake of calcium, vitamin d, and dairy products: a meta-analysis of 26,335 cases from 60 observational studies. Nutr Cancer. 2009;61(1):47-69.
51. Opotowsky AR, Bilezikian JP. Racial differences in the effect of early milk consumption on the peak and postmenopausal bone mineral density. J Bone Miner Res. 2003;8:1978-1988.
52. Ballew C, Kuester S, Gillespie C. Beverage choices affect adequacy of children's nutrient intakes. Arch Pediatr Adolesc Med. 2000;154:1148–1147.
53. Bowman SA. Beverage choices of young females: changes and impact on nutrient intakes. J Am Diet Assoc. 2002;102:1234–1239.
54. Phillips S, Jacobs Starkey L, Gray-Donald K.Food habits of Canadians: food sources of nutrients for the adolescent sample. Can J Diet Pract Res. 2004; 65:81-4.
55. Storey ML, Forshee RA, Anderson PA. Beverage consumption in the US population. J Am Diet Assoc. 2006; 106:1992-2000.
56. Hirsch AR, Förch K, Neidhart S, Wolf G, Carle R. Effects of thermal treatments and storage on pectin methylesterase and peroxidase activity in freshly squeezed orange juice. J Agric Food Chem. 2008 Jul 23;56(14):5691-9. Epub 2008 Jun 27.
57. Díaz-Juárez JA, Tenorio-López FA, Zarco-Olvera G, Valle-Mondragón LD, Torres-Narváez JC, Pastelín-Hernández G. Effect of Citrus paradisi extract and juice on arterial pressure both in vitro and in vivo. Phytother Res. 2009 Jan 19. [Epub ahead of print].
58. Di Martino P, Agniel R, David K, Templer C, Gaillard JL, Denys P, Botto H. Reduction of Escherichia coli adherence to uroepithelial bladder cells after consumption of cranberry juice: a double-blind randomized placebo-controlled cross-over trial. World J Urol. 2006 Feb;24(1):21-7. Epub 2006 Jan 6.

59. Zhang Y, Guo X, Saitz R, Levy D, Sartini E, Niu J, Ellison RC. Secular trends in alcohol consumption over 50 years: the Framingham Study. Am J Med. 2008 Aug;121(8):695-701.

60. Mukamal KJ, Rimm EB. Alcohol consumption: risks and benefits. Curr Atheroscler Rep. 2008 Dec;10(6):536-43. Review.

61. Panel on Dietary Reference Intakes for Electrolytes and Water, Standing Committee on the Scientific Evaluation of Dietary Reference Intakes, Food and Nutrition Board, Institute of Medicine (2004) Dietary Reference Intakes for Water, Potassium, Sodium, Chloride, and Sulfate. National Academy Press Washington, DC.

62. Popkin BM, Barclay DV, Nielsen SJ. Water and food consumption patterns of U.S. adults from 1999 to 2001. Obes Res 2005;13:2146-2152.

63. Soffritti M, Belpoggi F, Tibaldi E, Esposti DD, Lauriola M. Life-span exposure to low doses of aspartame beginning during prenatal life increases cancer effects in rats. Environ Health Perspect. 2007 Sep;115(9):1293-7.

64. Magnuson BA, Burdock GA, Doull J, Kroes RM, Marsh GM, Pariza MW, Spencer PS, Waddell WJ, Walker R, Williams GM. Aspartame: a safety evaluation based on current use levels, regulations, and toxicological and epidemiological studies. Crit Rev Toxicol. 2007;37(8):629-727.

65. Nawrot P, Jordan S, Eastwood J, Rotstein J, Hugenholtz A, Feeley M. Effects of caffeine on human health. Food Addit Contam. 2003 Jan;20(1):1-30. Review.

66. Elizabeth A. Dennis, Kyle D. Flack, Brenda M. Davy. Beverage consumption and adult weight management: A review. Eating Behaviors 10 (2009) 237–246.

67. Christine M. Bachman, Tom Baranowski, and Theresa A. Nicklas. Is There an Association Between Sweetened Beverages and Adiposity? International Life Sciences Institute doi: 10.1301/nr.2006.apr.153–174.

68. DiMeglio DP, Mattes RD. Liquid versus solid carbohydrate: effects on food intake and body weight. Int J Obes Relat Metab Disord. 2000;24:794–800.

69. Raben A, Vasilaras TH, Moller AC, Astrup A. Sucrose compared with artificial sweeteners: different effects on ad libitum food intake and body weight after 10 wk of supplementation in overweight subjects. Am J Clin Nutr. 2002;76:721–729.

70. De Castro JM. The effects of the spontaneous ingestion of particular foods or beverages on the meal pattern and overall nutrient intake of humans. Physiol Behav. 1993;53:1133-1144. Abstract.
71. St-Onge MP, Rubiano F, DeNino WF, et al. Added thermogenic and satiety effects of a mixed nutrient vs a sugar-only beverage. Int J Obes Relat Metab Disord. 2004;28:248-253. Abstract.
72. Brand-Miller JC, Holt SHA, Pawlak DB, McMillan J. Glycemic index and obesity. Am J Clin Nutr. 2002;76:281S–285S.
73. Ludwig D. Dietary Glycemic index and obesity. J Nutr. 2000;30:280S–283S.
74. Mattes RD. Dietary compensation by humans for supplemental energy provided as ethanol or carbohydrate in fluids. Physiol Behav. 1996;59:179– 187.
75. Zemel MB. Regulation of adiposity and obesity risk by dietary calcium: mechanisms and implications. J Am Coll Nutr. 2002;21:146S–151S.
76. Zemel MB. Role of calcium and dairy products in energy partitioning and weight management. Am J Clin Nutr. 2004;79:907S–912S.
77. French SA, Jeffery RW, Forster JL, McGovern PG, Kelder SH, Baxter JE. Predictors of weight change over two years among a population of working adults: the Healthy Worker Project. Int J Obes Relat Metab Disord. 1994;18:145–154.
78. Liebman M, Pelican S, Moore SA, et al. Dietary intake, eating behavior, and physical activity-related determinants of high body mass index in rural communities in Wyoming, Montana, and Idaho. Int J Obes Relat Metab Disord. 2003;27:684–692.
79. Lin BH, Huang CL, French SA. Factors associated with women's and children's body mass indices by income status. International Journal of Obesity and Related Metabolic Disorders. 2004; 28(4), 536−542.
80. McCarthy SN, Robson PJ, Livingstone MB, et al. Associations between daily food intake and excess adiposity in Irish adults: towards the development of food-based dietary guidelines for reducing the prevalence of overweight and obesity. Int J Obes (Lond). 2006; 30, 993–1002.
81. Schulz M, Kroke A, Liese AD, Hoffmann K, Bergmann MM, Boeing H. Food groups as predictors for short-term weight changes in men and women of the EPIC-Potsdam cohort. J Nutr. 2002; 132: 1335–1340.

82. Newby PK, Muller D, Hallfrisch J, Qiao N, Andres R, Tucker KL. Dietary patterns and changes in body mass index and waist circumference in adults. Am J Clin Nutr. 2003;77:1417-25.

83. Schulze MB, Manson JE, Ludwig DS, et al. Sugar-sweetened beverages, weight gain, and incidence of type 2 diabetes in young and middle-aged women. JAMA. 2004; 292: 927–34.

84. Nooyens AC, Visscher TL, Schuit AJ, et al. Effects of retirement on lifestyle in relation to changes in weight and waist circumference in Dutch men: a prospective study. Public Health Nutr.2005; 8, 1266–1274.

85. Bes-Rastrollo M, Sanchez-Villegas A, Gomez-Gracia E, Martinez JA, Pajares RM, Martinez-Gonzalez MA. Predictors of weight gain in a Mediterranean cohort: the Seguimiento Universidad de Navarra Study 1. Am J Clin Nutr. 2006; 83: 362–70.

86. Dhingra R, Sullivan L, Jacques PF,Wang TJ, Fox CS, Meigs JB, et al. Soft drink consumption and risk of developing cardiometabolic risk actors and the metabolic syndrome in middle-aged adults in the community. Circulation. 2007; 116(5), 480−488.

87. Tordoff MG, Alleva AM. Effect of drinking soda sweetened with aspartame or high-fructose corn syrup on food intake and body weight. American Journal of Clinical Nutrition1990; 51(6), 963−969.

88. Chen L, Appel LJ, Loria C, Lin PH, Champagne CM, Elmer PJ, Ard JD, Mitchell D, Batch BC, Svetkey LP, Caballero B. Reduction in consumption of sugar-sweetened beverages is associated with weight loss: the PREMIER trial. Am J Clin Nutr 2009;89:1299–306.

89. Vartanian LR, Schwartz MB, Brownell KD. Effects of soft drink consumption on nutrition and health: A systematic review and meta- nalysis. American Journal of Public Health 2007;97(4), 667‒675.

90. Davies KM, Heaney RP, Recker RR, Lappe JM, Barger-Lux MJ, et al. Calcium Intake and body weight. J Clin Endocrinol Metab 2000;85:4635e8.

91. Parikh SJ, Yanovski JA. Calcium intake and adiposity. Am J Clin Nutr 2003;77:281e7. and Pereira MA, Jacobs Jr DR, Van Horn L, Slattery ML, Kartashov AI, Ludwig DS. Dairy consumption, obesity, and the insulin resistance syndrome in young adults: the CARDIA Study. JAMA 2002;287:2081e9.

92. Rosell M, Johansson G, Berglund L, Vessby B, de Faire U, Hellenius ML. Associations between the intake of dairy fat and calcium and abdominal obesity. Int J Obes Relat Metab Disord 2004;28:1427e34.
93. Loos RJ, Rankinen T, Leon AS, Skinner JS, Wilmore JH, Rao DC, et al. Calcium intake is associated with adiposity in Black and White men and White women of the HERITAGE Family Study. J Nutr 2004;134:1772e8.
94. Marques-Vidal P, Goncalves A, Dias CM. Milk intake is inversely related to obesity in men and in young women: data from the Portuguese Health Interview Survey 1998e 1999. Int J Obes 2006;30:88e93.
95. Rajpathak SN, Rimm EB, Rosener B, Willett WC, Hu FB. Calcium and dairy intakes in relation to long-term weight gain in US men. Am J Clin Nutr 2006;83:559–66.
96. Drapeau V, Despres J-P, Bouchard C, Allard L, Fournier G, Leblanc C, et al. Modifications in food-group consumption are related to long-term body weight changes. Am J Clin Nutr 2004;80:29e37.
97. Zemel MB, Thompson W, Milstead A, Morris K, Campbell P. Calcium and dairy acceleration of weight and fat loss during energy restriction in obese adults. Obes Res 2004; 12:582e90.
98. Zemel MB, Richards J, Mathis S, Milstead A, Gebhardt L, Silva E. Dairy augmentation of total and central fat loss in obese subjects. Int J Obes 2005;29:391e7.
99. Thompson WG, Holdman NR, Janzow DJ, Slezak JM, Morris KL, Zemel MB. Effects of energy-reduced diets high in dairy products and fiber on weight loss in obese adults. Obes Res 2005;13:1344e53.
100. Harvey-Berino J, Casey Gold B, Lauber R, Starinski A. The impact of calcium and dairy product consumption on weight loss. Obes Res 2005;13:1720e6.
101. Lanou AJ, Barnard ND. Dairy and weight loss hypothesis: an evaluation of the clinical trials. Nutr Rev. 2008 May;66(5):272-9.
102. Jacobsen R, Lorenzen JK, Toubro S, Krog-Mikkelsen I, Astrup A. Effect of short-term high dietary calcium intake on 24-h energy expenditure, fat oxidation, and fecal fat excretion. Int J Obes Relat Metab Disord 2005, 29:292-301.
103. Papakonstantinou E, Flatt WP, Huth PJ, Harris RB. High dietary calcium reduces body fat content, digestibility of fat, and serum vitamin D in rats. Obes Res 2003;11:387e94.

104. Lukaszuk JM, Luebbers P, Gordon BA. Preliminary study: soy milk as effective as skim milk in promoting weight loss. J Am Diet Assoc. 2007 Oct;107(10):1811-4.
105. Westphal S, Kastner S, Taneva E, Leodolter A, Dierkes J, Luley K. Postprandial lipid and carbohydrates responses after the ingestion of a casein-enriched mixed meal. Am J Clin Nutr 2004;80:284e90.
106. Sun X, Zemel MB: Leucine and calcium regulate fat metabolism and energy partitioning in murine adipocytes and muscle cells. Lipids 2007, 42:297-305.
107. Pihlanto-Leppala A, Koskinen P, Piilola K, Tupasela T, Korhonen H. Angiotensin I-converting enzyme inhibitory properties of whey protein digests: concentration and characterization of active peptides. J Dairy Res 2000;67: 53e64.
108. Alonso A, Zozaya C, Vázquez Z, Alfredo Martínez J, Martínez-González MA. The effect of low-fat versus whole-fat dairy product intake on blood pressure and weight in young normotensive adults. J Hum Nutr Diet. 2009 Aug;22(4):336-42. Epub 2009 May 26.
109. Nicklas TA, O'Neil CE, Kleinman R. Association between 100% juice consumption and nutrient intake and weight of children aged 2 to 11 years. Arch Pediatr Adolesc Med. 2008 Jun;162(6):557-65.
110. Block G, Dresser CM, Hartman AM, Carroll MD. Nutrient sources in the American diet: quantitative data from the NHANES II survey. Am J Epidemiol. 1985;122:27–40.
111. Suter PM. Is alcohol consumption a risk factor for weight gain and obesity? Crit Rev Clin Lab Sci. 2005;42(3):197-227.
112. Sakurai Y, Umeda T, Shinchi K, Honjo S, Wakabayashi K, Todoroki I, Nishikawa H, Ogawa S, Katsurada M. Relation of total and beverage-specific alcohol intake to body mass index and waistto- hip ratio: a study of self-defense officials in Japan. Eur J Epidemiol 1997; 13: 893–898.
113. Riserus, U., & Ingelsson, E. (2007). Alcohol intake, insulin resistance, and abdominal obesity in elderly men. Obesity (Silver Spring), 15(7), 1766−1773.
114. Clevidence BA, Taylor PR, Campbell WS, Judd JT. Lean and heavy women may not use energy from alcohol with equal efficiency. J Nutr 1995; 125: 2536–2540.
115. Colditz GA, Giovannucci E, Rimm EB, Stampfer MJ, Rosner B, Speizer FE, Gordis E, Willett WC. Alcohol intake in relation to diet and obesity in women and in men. Am J Clin Nutr 1991; 54: 49– 55.

116. Lahti-Koski M, Pietinen P, Heliövaara M, Vartiainen E.Associations of body mass index and obesity with physical activity, food choices, alcohol intake, and smoking in the 1982-1997 FINRISK Studies. Am J Clin Nutr. 2002 May;75(5):809-17.

117. Appleby PN, Thorogood M, Mann JI, Key TJ. Low body mass index in non-meat eaters: the possible roles of animal fat, dietary fibre and alcohol. Int J Obes Relat Metab Disord 1998; 22: 454– 460.

118. Suter PM, H¨asler E, Meyer R, Vetter W. Modulation of the alcohol effects on body weight by the family history of overweight. Int J Obesity 1999; 23(Suppl 5): S94.

119. Sepp¨aK, Sillanaukee P, Pitk¨aj¨arviT, Nikkil¨aM, KoivulaT. Moderate and heavy alcohol consumption have no favorable effect on lipid values. Arch Intern Med 1992; 152: 297–300.

120. Fonager K, Sabroe S. A comparative analysis of different methods for obtaining estimates of alcohol consumption in a Danish population survey. Scand J Public Health. 2001;29(4):256-62.

121. Stookey JD, Constant F, Popkin BM, Gardner CD. Drinking water is associated with weight loss in overweight dieting women independent of diet and activity. Obesity (Silver Spring). 2008 Nov;16(11):2481-8. Epub 2008 Sep 11.

122. Stookey JD, Constant F, Gardner CD, Popkin BM. Replacing sweetened caloric beverages with drinking water is associated with lower total energy intake. Obes Res 2007;15:3013–3022.

123. Greenberg JA, Boozer CN, Geliebter A. Coffee, diabetes, and weight control. Am J Clin Nutr. 2006 Oct;84(4):682-93. Review.

124. Lopez-Garcia E, van Dam RM, Rajpathak S, Willett WC, Manson JE, Hu FB. Changes in caffeine intake and long-term weight change in men and women. Am J Clin Nutr. 2006 Mar;83(3):674-80.

125. Fowler SP, Williams K, Resendez RG, Hunt KJ, Hazuda HP, Stern MP. Fueling the obesity epidemic? Artificially sweetened beverage use and long-term weight gain. Obesity (Silver Spring). 2008 Aug;16(8):1894-900. Epub 2008 Jun 5.

126. Martín AR, Nieto JM, Ruiz JP, Jiménez LE. Overweight and obesity: the role of education, employment and income in Spanish adults. Appetite. 2008 Sep;51(2):266-72. Epub 2008 Mar 7.

127. Tjepkema M. Adult obesity. Health Reports. 2006 Aug;17(3):9-25.

128. Hanson KL, Sobal J, Frongillo EA. Gender and marital status clarify associations between food insecurity and body weight. J Nutr. 2007 Jun;137(6):1460-5.

129. Batnitzky A. Obesity and household roles: gender and social class in Morocco. Sociol Health Illn. 2008 Apr;30(3):445-62. Epub 2008 Mar 26.

130. Siervo M, Davies AA, Jebb SA, Jalil F, Moore SE, Prentice AM. Ethnic differences in the association between body mass index and impedance index (Ht2/Z) in adult women and men using a leg-to-leg bioimpedance method. Eur J Clin Nutr. 2007 Nov;61(11):1337-40. Epub 2007 Feb 14.

131. Gillum RF, Sempos CT. Ethnic variation in validity of classification of overweight and obesity using self-reported weight and height in American women and men: the Third National Health and Nutrition Examination Survey. Nutr J. 2005 Oct 6;4:27.

132. Arambepola C, Allender S, Ekanayake R, Fernando D. Urban living and obesity: is it independent of its population and lifestyle characteristics? Trop Med Int Health. 2008 Apr;13(4):448-57. Epub 2008 Mar 10.

133. Nyholm M, Gullberg B, Haglund B, Råstam L, Lindblad U. Higher education and more physical activity limit the development of obesity in a Swedish rural population. The Skaraborg Project. Int J Obes (Lond). 2008 Mar;32(3):533-40. Epub 2008 Jan 22.

134. Shields M, Tjepkema M. Regional differences in obesity. Health Reports. 2006 Aug;17(3):61-7.

135. Sussner KM, Lindsay AC, Greaney ML, Peterson KE. The influence of immigrant status and acculturation on the development of overweight in Latino families: a qualitative study. J Immigr Minor Health. 2008 Dec;10(6):497-505.

136. Bates LM, Acevedo-Garcia D, Alegría M, Krieger N. Immigration and generational trends in body mass index and obesity in the United States: results of the National Latino and Asian American Survey, 2002-2003. Am J Public Health. 2008 Jan;98(1):70-7. Epub 2007 Nov 29.

137. Sobal J, Rauschenbach BS, Frongillo EA Jr. Marital status, fatness and obesity. Soc Sci Med. 1992 Oct;35(7):915-23.

138. Tjepkema M. Measured obesity, Adult obesity in Canada:Measured height and weight. 2005. Component of Statistics Canada Catalogue no. 82-620, page 1-32.

139. Norman A, Bellocco R, Vaida F, Wolk A. Total physical activity in relation to age, body mass, health and other factors in a cohort of Swedish men. Int J Obes Relat Metab Disord. 2002 May;26(5):670-5.

140. Drewnowski A, Specter SE. Poverty and obesity: the role of energy density and energy costs. Am J Clin Nutr. 2004 Jan;79(1):6-16.

141. Drewnowski A, Monsivais P, Maillot M, Darmon N. Low-energy-density diets are associated with higher diet quality and higher diet costs in French adults. J Am Diet Assoc. 2007 Jun;107(6):1028-32.

142. Drewnowski A, Darmon N. The economics of obesity: dietary energy density and energy cost. Am J Clin Nutr. 2005 Jul;82(1 Suppl):265S-273S.

143. Hill JO, Melanson EL. Overview of the determinants of overweight and obesity: current evidence and research issues. Med Sci Sports Exerc. 1999 Nov;31(11 Suppl):S515-21.

144. Weinsier RL, Hunter GR, Heini AF, Goran MI, Sell SM. The etiology of obesity: relative contribution of metabolic factors, diet, and physical activity. Am J Med. 1998 Aug;105(2):145-50.

145. Saarni SE, Pietiläinen K, Kantonen S, Rissanen A, Kaprio J. Association of smoking in adolescence with abdominal obesity in adulthood: a follow-up study of 5 birth cohorts of Finnish twins. Am J Public Health. 2009 Feb;99(2):348-54. Epub 2008 Dec 4.

146. Padrão P, Lunet N, Santos AC, Barros H. Smoking, alcohol, and dietary choices: evidence from the Portuguese National Health Survey. BMC Public Health. 2007 Jul 3;7:138.

147. Maureen LS, Richard AF, Patricia AA. Beverage Consumption in the US Population. J Am Diet Assoc. 2006;106:1992-2000.

148. de Castro JM, Taylor T. Smoking status relationships with the food and fluid intakes of free-living humans. Nutrition. 2008 Feb;24(2):109-19.

149. Rehm CD, Matte TD, Van Wye G, Young C, Frieden TR. Demographic and behavioral factors associated with daily sugar-sweetened soda consumption in New York City adults. J Urban Health. 2008 May;85(3):375-85. Epub 2008 Mar 18.

150. Kvaavik E, Andersen LF, Klepp KI. The stability of soft drinks intake from adolescence to adult age and the association between long-term consumption of soft drinks and lifestyle factors and body weight. Public Health Nutr. 2005 Apr;8(2):149-57.

151. Capita R, Alonso-Calleja C. Differences in reported winter and summer dietary intakes in young adults in Spain. Int J Food Sci Nutr. 2005 Sep;56(6):431-43.

152. Canadian Community Health Survey- Nutrition (CCHS 2.2). Detailed information for 2004, definitions, data sources, and methods. Statistics Canada. Available at URL: http://www.statcan.gc.ca/cgi-bin/imdb/p2SV.pl?Function=getSurvey&SDDS=5049&lang=en&db=imdb&adm=8&dis=2

153. Tucker KL. Dietary patterns, approaches, and multicultural perspective. Appl. Physiol.Nutr. Metab. 2010; 35: 211-218.

154. Moeller SM, Reedy J, Millen AE, Dixon LB, Newby PK, Tucker KL, Krebs-Smith SM, Guenther PM. Dietary patterns: challenges and opportunities in dietary patterns research an Experimental Biology workshop, April 1, 2006. J Am Diet Assoc. 2007 Jul;107(7):1233-9.

155. LaRowe TL, Moeller SM, Adams AK. Beverage patterns, diet quality, and body mass index of US preschool and school-aged children. J Am Diet Assoc. 2007; 107:1124-1133.

156. Duffey KJ, Popkin BM. Shifts in patterns and consumption of beverages between 1965 and 2002. Obesity. 2007; 15:2739-47.

157. Wirfält AKE, Jeffery RW. Using Cluster Analysis to Examine Dietary Patterns: Nutrient Intakes, Gender, and Weight Status Differ Across Food Pattern Clusters. Journal of American Dietetic Association. 1997; 97:272-279.

158. James DCS. Cluster analysis defines distinct dietary patterns for African-American men and women. J Am Diet Assoc. 2009;109: 255-262.

159. Canadian Community Health Survey (CCHS) Cycle 2.2 (2004). Nutrition-General health (including vitamins & mineral supplements) & 24-hour dietary recall components, user guide. April 2008. Available at URL: http://www.statcan.gc.ca/imdb-bmdi/document/5049_D24_T9_V1-eng.pdf

160. Shields M, Gorber SC, Tremblay MS. Estimates of obesity based on self-report versus direct measures. Health Reports (statistics Canada, Catalogue 82-003) 2008; 19(2): 61-76.

161. Michels KB, Willett WC. Self-administered semiquantitative food frequency questionnaires: patterns, predictors, and interpretation of omitted items. Epidemiology. 2009 Mar;20(2):295-301.

162. Witte JS, Ursin G, Siemiatycki J, Thompson WD, Paganini-Hill A, Haile RW. Diet and premenopausal bilateral breast cancer: a case-control study. Breast Cancer Res Treat. 1997 Feb;42(3):243-51.
163. Garriguet D. Impact of identifying plausible respondents on the under-reporting of energy intake in the Canadian Community Health Survey. October 2008. Component of Statistics Canada no. 82-003-X.
164. Institute of Medicine. Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein and Amino Acids. Washington, DC: National Academy Press, 2005.
165. SAS User Guide. The FASTCLUS procedure. Available at URL: http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#/documentation/cdl/en/statug/63033/HTML/default/fastclus_toc.htm and http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#/documentation/cdl/en/statug/63033/HTML/default/statug_cluster_sect002.htm
166. Pahwa P. Biostatistics one [course materials]. University of Saskatchewan, Department of Community Health and Epidemiology; 2007.
167. Malik VS, Popkin BM, Bray GA, Després JP, Willett WC, Hu FB. Sugar Sweetened Beverages and Risk of Metabolic Syndrome and Type 2 Diabetes: A Meta-analysis. Diabetes Care. 2010 Aug 6.
168. National Health and Nutrition Examination Survey (NHANES), Introduction to NHANES. Available at URL: http://www.cdc.gov/nchs/nhanes/about_nhanes.htm
169. Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999-2004. JAMA. 2006 Apr 5;295(13):1549-55.
170. Garriguet D. Canadians' eating habits. Health Rep. 2007 May;18(2):17-32.
171. Statistics Canada. Food statistics 2009 No. 21.020-X). Ottawa: Statistics Canada.
172. Beridot-Therond ME, Arts I, Fantino M, de la Gueronniere V. Short-term effects of the flavour of drinks on ingestive behaviours in man. Appetite. 1998;31:67-81.
173. Rolls BJ, Kim S, Fedoroff IC. Effects of drinks sweetened with sucrose or aspartame on hunger, thirst and food intake in men. Physiol Behav. 1990;48:19-26.

174. De Castro JM. The effects of the spontaneous ingestion of particular foods or beverages on the meal pattern and overall nutrient intake of humans. Physiol Behav. 1993;53:1133-1144.

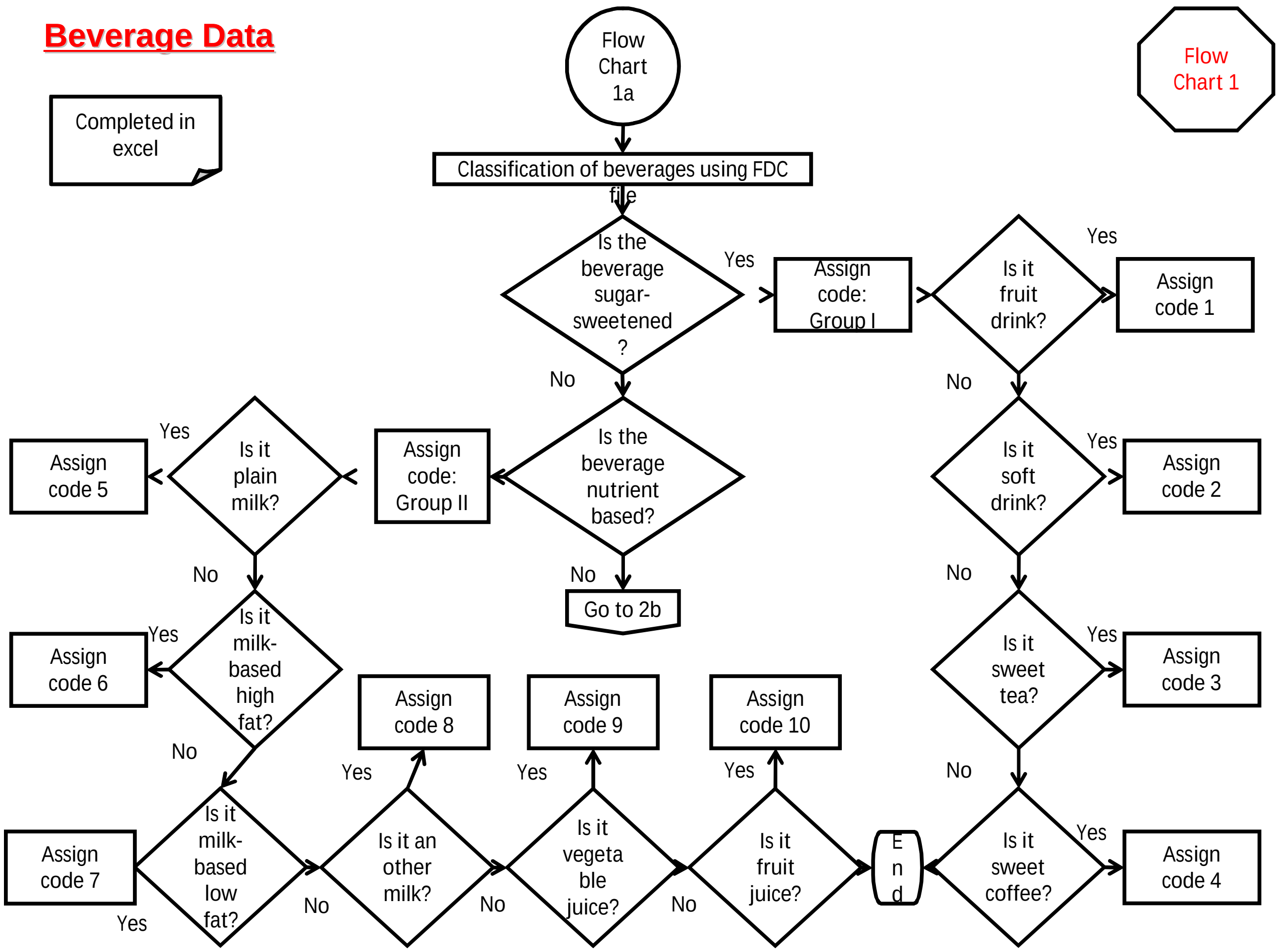
175. Sclafani A. Starch and sugar tastes in rodents: an update. Brain Res Bull. 1991;27:383-386.

APPENDICES

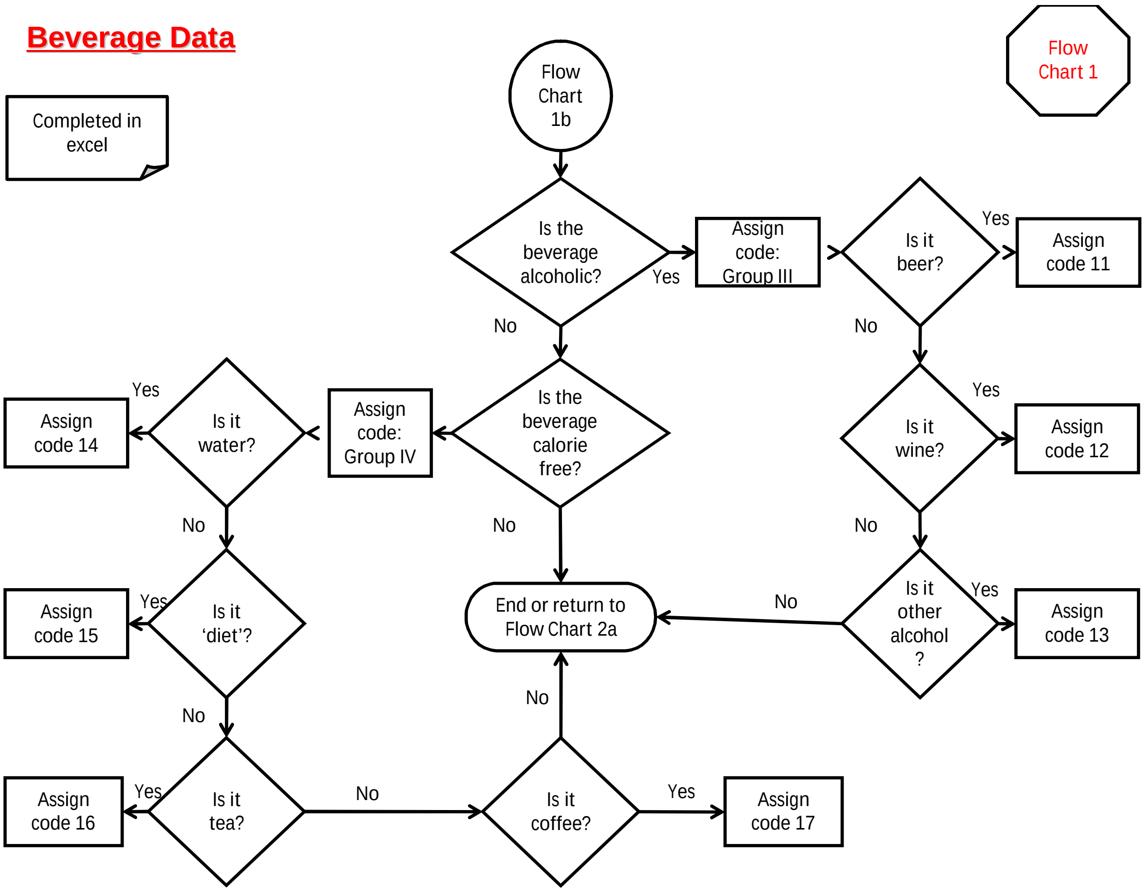
Appendix A: Ethics Approval

Since this study involved analyses of secondary data, the participants were not deemed to be at any risk. Specifically, because the data was already de-identified, the researchers were not able to associate specific information with specific participants, and there was no risk for deception, loss of confidentiality, or anonymity. In data analyses, the Statistics Canada Research Data Center guidelines were followed to avoid small cells sizes. The data were reported in aggregate results. These measures ensured confidentiality and anonymity of CCHS 2.2 participants.

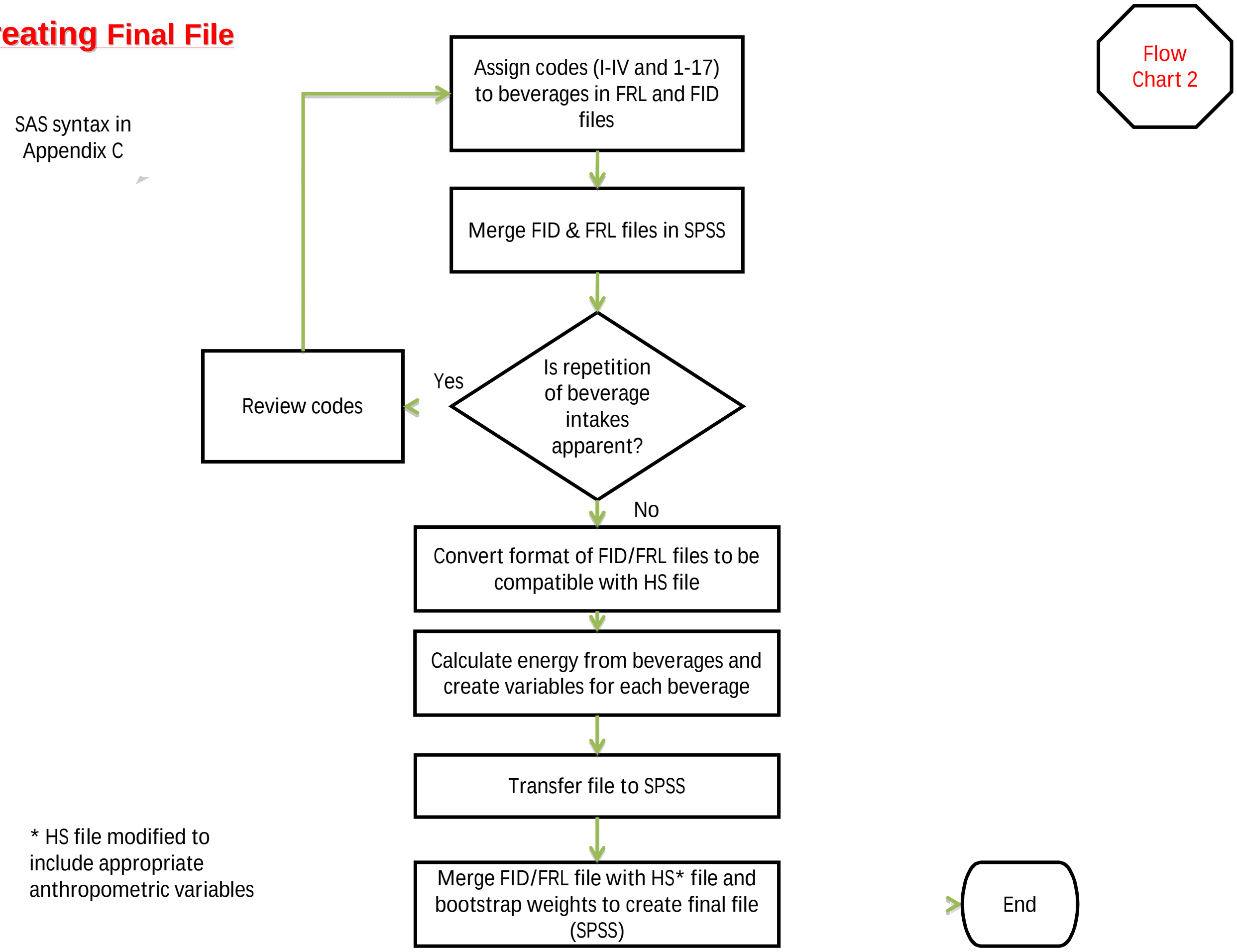
Appendix B: Flow Charts



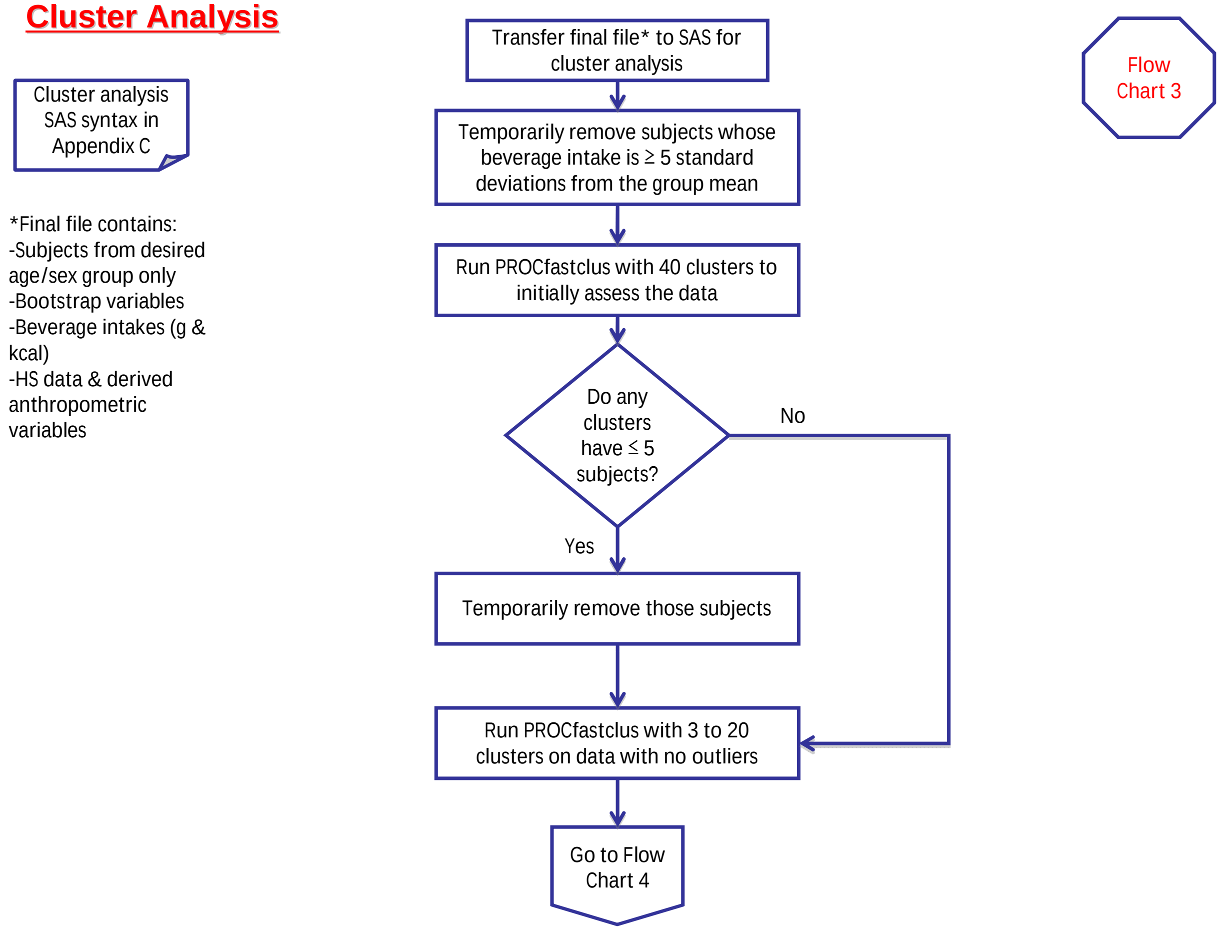
Beverage Data



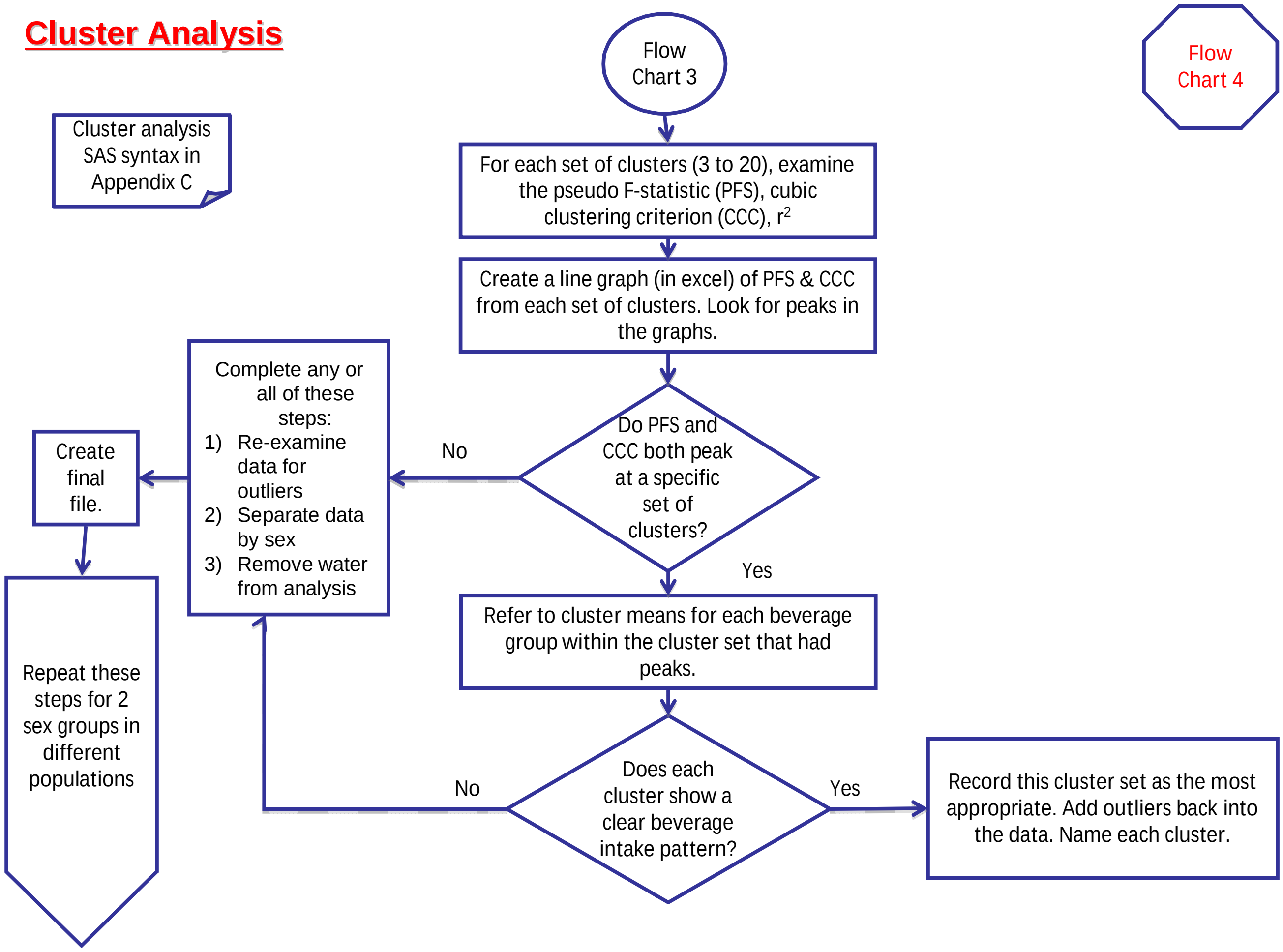
Creating Final File



Cluster Analysis



Cluster Analysis



Appendix C: Syntaxes

Recoding beverage groups:

```
libname dir "P:\07-SSH-PRC-1331\Nooshin";
data dir.fid_new;
set  "P:\07-SSH-PRC-1331\Nooshin\SAS\English\Foods, Ingredients and
Recipes\fid_dm.sas7bdat";
run;
data dir.fid_new;
set dir.fid_new;
newcode=0;

if FIDD_CDE in
('2885','2889','2891','2893','2895','2904','2917','2922','2940','2954',
'2955','2956','2958','2959','2960','2961','2965','2967','2968','2969',
'2972','2974','2976','2981','2983','4984','5287','1725','12774','12775',
'12776','12777','12778','12779','12780','12781','12782','12783','12784',
'12785','12786','12787','12788','12789','12790','12791','12792','12793',
'12794','12795','12796','12797','12798','12799','12800','12801','12802',
'12803','13185','13192','13513','500609','500737','500738','501137','501162',
'501163','501164','13186') then newcode=1;

else if FIDD_CDE in
('5293','2854','2855','2856','2857','2858','2859','2860','2861','2966','2920',
'4980','5288','501148') then newcode=2;

else if FIDD_CDE in
('4908','12764','12765','12768','12772','12773','13196','2915','501150')
then newcode=3;

else if FIDD_CDE in
('2928','2929','2930','12751','12752','12753','13180','13181','13182')
then newcode=4;

else if FIDD_CDE in
('61','62','63','64','65','114','10000','10007','10009','10012','10014',
'10016','501090','500585','500586') then newcode=5;

else if FIDD_CDE in
('55','69','75','76','113','116','123','2863','2865','2867','2871','2887',
'2897','2899','2901','2903','2905','2906','2908','2931','2932','2970','5295',
'5296','5297','10025','10028','10032','10033','10034','10036','10037','10038',
'10039','10041','10042','10045','10046','10047','10048','10070','10071',
'13176','13177','13184','13188','13189','13190','13191','13210','13211',
'13212','13213','500844','500916','500917','500918','501092','501132','13171',
'13172','13173') then newcode=6;

else if FIDD_CDE in
('70','71','4711','5268','5269','5270','5271','5272','5273','5274','5275',
'5276','10023','10026','10027','10029','10030','10031','10035','10040',
'10049','12842','12843','13193','13197','13287','13288','10024')
```

```
then newcode=7;

else if FIDD_CDE in
('72','73','74','77','79','124','3402','4780','5241','500669')then newcode=8;

else if FIDD_CDE in
('2312','2464','2473','2868','500658')then newcode=9;

else if FIDD_CDE in
('1485','1495','1497','1570','1572','1576','1589','1590','1591','1594','1595',
'1619','1620','1622','1624','1625','1627','1629','1631','1632','1644','1652','
1657','1659','1673','1694','1716','1717','1720','1723','1752','1754',
'12085','12086','12087','12088','12090','12091','12092','12094','12095',
'12096','12097','12098','12099','12100','12101','12102','12103','12104',
'12105','12174','12175','12176','12177','12178','12179','12180','12181',
'12182','12183','12184','12185','12186','12187','12188','12189','500690',
'500691','500934','500935','501159','501161') then newcode=10;

else if FIDD_CDE in
('2835','2943','13501')then newcode=11;

else if FIDD_CDE in
('501096','2848','2849','2850','2851','2852','2937') then newcode=12;

else if FIDD_CDE in
('2837','2838','2842','2843','2844','2845','2846','2847','2924','2925','2934',
'2935','2936','2948','2949','2951','2952','2977','2978','12804','12805',
'12806','12807','12808','12809','12810','12811','12812','12813','12814',
'12815','12816','12817','12818','12819','12820','12821','12822','12823',
'12824','12825','12826','12827','12828','12829','12830','12831','12833',
'12834','12835','12836','12837','12838','12839','12840','12841','13169',
'13170','500971') then newcode=13;

else if FIDD_CDE in
('2853','2918','2919','2933','13465')then newcode=14;


else if FIDD_CDE in
('2853','2926','2938','2963','5289','5292','13187','13470','13493','13546',
'5294')then newcode=15;

else if FIDD_CDE in
('2909','2911','2913','2916','2941','2942','4985','12762','12763','12770',
'13194','13195')then newcode=16;

else if FIDD_CDE in
('2872','2873','2875','2877','2879','2884','4779','12745','12746','12747',
'12748','12750','12754','12755','12756','12757','12758','12759','12760',
'12761','13179','500843','500845')then newcode=17;


else newcode=18;
run;
proc freq data=dir.fid_new;
tables newcode;
run;
```

Syntax to make the beverage file compatible with HS file (Energy):

```
libname dir "P:\07-SSH-PRC-1331\Nooshin\Aug25\New Folder";
data temp;
set dir.small_fid_frl;
run;
proc contents;run;
proc print;run;
proc freq; tables newcode;run;

data
frdnk sfdnk tswt cofswt milk mlk_hi mlk_lw mlk_oth veg_ju fr_ju beer wine
alc_othr water diet tea coffee;
set temp;

if newcode=1 then output frdnk ;
if newcode=2 then output sfdnk ;
if newcode=3 then output tswt ;
if newcode=4 then output cofswt ;
if newcode=5 then output milk ;
if newcode=6 then output mlk_hi ;
if newcode=7 then output mlk_lw ;
if newcode=8 then output mlk_oth;
if newcode=9 then output veg_ju ;
if newcode=10 then output fr_ju ;
if newcode=11 then output beer ;
if newcode=12 then output wine ;
if newcode=13 then output alc_othr;
if newcode=14 then output water ;
if newcode=15 then output diet ;
if newcode=16 then output tea ;
if newcode=17 then output coffee ;
run;

* Fruits drinks;
proc sort data = frdnk ; by sampleid;
run;
proc means data= frdnk noprint;
by sampleid;
var fidd_ekc;
output out=sum_frdnk sum=fidd_ekc;
run;
data sum_frdnk;
set sum_frdnk;
frdn_cal=fidd_ekc;
run;

data sum_frdnk;
set sum_frdnk;
keep sampleid frdn_cal;
run;
proc sort data=sum_frdnk;
by sampleid;
run;

* soft  drinks;
```

```
proc sort data = sfdnk ; by sampleid;
run;
proc means data= sfdnk noprint;
by sampleid;
var fidd_ekc;
output out=sum_sfdnk sum=fidd_ekc;
run;
data sum_sfdnk;
set sum_sfdnk;
sfdn_cal=fidd_ekc;
run;

data sum_sfdnk;
set sum_sfdnk;
keep sampleid sfdn_cal;
run;
proc sort data=sum_sfdnk;
by sampleid;
run;


* tea-sweetened;
proc sort data = tswt ; by sampleid;
run;
proc means data= tswt noprint;
by sampleid;
var fidd_ekc;
output out=sum_tswt sum=fidd_ekc;
run;
data sum_tswt;
set sum_tswt;
tswt_cal=fidd_ekc;
run;

data sum_tswt;
set sum_tswt;
keep sampleid tswt_cal;
run;
proc sort data=sum_tswt;
by sampleid;
run;


*coffee_sweetened;
proc sort data = cofswt ; by sampleid;
run;
proc means data= cofswt noprint;
by sampleid;
var fidd_ekc;
output out=sum_cfswt sum=fidd_ekc;
run;
data sum_cfswt;
set sum_cfswt;
cfswt_cal=fidd_ekc;
run;
```

```
data sum_cfswt;
set sum_cfswt;
keep sampleid cfswt_cal;
run;
proc sort data=sum_cfswt;
by sampleid;
run;

*plain milk;
proc sort data = milk ; by sampleid;
run;
proc means data= milk noprint;
by sampleid;
var fidd_ekc;
output out=sum_milk sum=fidd_ekc;
run;
data sum_milk;
set sum_milk;
milk_cal=fidd_ekc;
run;

data sum_milk;
set sum_milk;
keep sampleid milk_cal;
run;
proc sort data=sum_milk;
by sampleid;
run;


*milk-based high-fat;
proc sort data = mlk_hi ; by sampleid;
run;
proc means data= mlk_hi noprint;
by sampleid;
var fidd_ekc;
output out=sum_mlkh sum=fidd_ekc;
run;
data sum_mlkh;
set sum_mlkh;
mlkh_cal=fidd_ekc;
run;

data sum_mlkh;
set sum_mlkh;
keep sampleid mlkh_cal;
run;
proc sort data=sum_mlkh;
by sampleid;
run;

*milk_based low_fat;
proc sort data = mlk_lw  ; by sampleid;
run;
proc means data= mlk_lw noprint;
by sampleid;
var fidd_ekc;
```

```
output out=sum_mlkl sum=fidd_ekc;
run;
data sum_mlkl;
set sum_mlkl;
mlkl_cal=fidd_ekc;
run;

data sum_mlkl;
set sum_mlkl;
keep sampleid mlkl_cal;
run;
proc sort data=sum_mlkl;
by sampleid;
run;

*other types of milk;

proc sort data = mlk_oth  ; by sampleid;
run;
proc means data= mlk_oth noprint;
by sampleid;
var fidd_ekc;
output out=sum_mlko sum=fidd_ekc;
run;
data sum_mlko;
set sum_mlko;
mlko_cal=fidd_ekc;
run;

data sum_mlko;
set sum_mlko;
keep sampleid mlko_cal;
run;
proc sort data=sum_mlko;
by sampleid;
run;

*vegatable juice;
proc sort data = veg_ju  ; by sampleid;
run;
proc means data= veg_ju noprint;
by sampleid;
var fidd_ekc;
output out=sum_vegj sum=fidd_ekc;
run;
data sum_vegj;
set sum_vegj;
vegj_cal=fidd_ekc;
run;

data sum_vegj;
set sum_vegj;
keep sampleid vegj_cal;
run;
proc sort data=sum_vegj;
by sampleid;
run;
```

```
*fruit juice;
proc sort data = fr_ju  ; by sampleid;
run;
proc means data= fr_ju noprint;
by sampleid;
var fidd_ekc;
output out=sum_frj sum=fidd_ekc;
run;
data sum_frj;
set sum_frj;
frj_cal=fidd_ekc;
run;

data sum_frj;
set sum_frj;
keep sampleid frj_cal;
run;
proc sort data=sum_frj;
by sampleid;
run;

*beer;
proc sort data = beer  ; by sampleid;
run;
proc means data= beer noprint;
by sampleid;
var fidd_ekc;
output out=sum_beer  sum=fidd_ekc;
run;
data sum_beer ;
set sum_beer ;
beer_cal=fidd_ekc;
run;

data sum_beer;
set sum_beer;
keep sampleid beer_cal;
run;
proc sort data=sum_beer;
by sampleid;
run;

*wine;
proc sort data = wine  ; by sampleid;
run;
proc means data= wine  noprint;
by sampleid;
var fidd_ekc;
output out=sum_wine   sum=fidd_ekc;
run;
data sum_wine  ;
set sum_wine  ;
wine_cal=fidd_ekc;
run;

data sum_wine ;
```

```
set sum_wine ;
keep sampleid wine_cal;
run;
proc sort data=sum_wine ;
by sampleid;
run;

*other alcoholic beverages;
proc sort data = alc_othr  ; by sampleid;
run;
proc means data= alc_othr  noprint;
by sampleid;
var fidd_ekc;
output out=sum_alcot   sum=fidd_ekc;
run;
data sum_alcot  ;
set sum_alcot  ;
alcot_cal=fidd_ekc;
run;

data sum_alcot ;
set sum_alcot ;
keep sampleid alcot_cal;
run;
proc sort data=sum_alcot ;
by sampleid;
run;

*water;
proc sort data = water  ; by sampleid;
run;
proc means data= water   noprint;
by sampleid;
var fidd_ekc;
output out=sum_water    sum=fidd_ekc;
run;
data sum_water   ;
set sum_water   ;
water_cal=fidd_ekc;
run;

data sum_water  ;
set sum_water  ;
keep sampleid water_cal;
run;
proc sort data=sum_water  ;
by sampleid;
run;

*diet drinks;
proc sort data = diet  ; by sampleid;
run;
proc means data= diet   noprint;
by sampleid;
var fidd_ekc;
output out=sum_diet    sum=fidd_ekc;
run;
```

```
data sum_diet   ;
set sum_diet   ;
diet_cal=fidd_ekc;
run;

data sum_diet  ;
set sum_diet  ;
keep sampleid diet_cal;
run;
proc sort data=sum_diet  ;
by sampleid;
run;

*tea;
proc sort data = tea  ; by sampleid;
run;
proc means data= tea   noprint;
by sampleid;
var fidd_wtg;
output out=sum_tea    sum=fidd_ekc;
run;
data sum_tea   ;
set sum_tea   ;
tea_cal=fidd_ekc;
run;

data sum_tea  ;
set sum_tea  ;
keep sampleid tea_cal;
run;
proc sort data=sum_tea  ;
by sampleid;
run;

*coffee;
proc sort data = coffee ; by sampleid;
run;
proc means data= coffee   noprint;
by sampleid;
var fidd_ekc;
output out=sum_coffee    sum=fidd_ekc;
run;
data sum_coffee   ;
set sum_coffee  ;
coffee_cal=fidd_ekc;
run;

data sum_coffee  ;
set sum_coffee  ;
keep sampleid coffee_cal;
run;
proc sort data=sum_coffee ;
by sampleid;
run;


data merg_bvg_oct;
```

```
merge  sum_frdnk sum_sfdnk sum_tswt sum_cfswt sum_milk sum_mlkh sum_mlkl
sum_mlko sum_vegj sum_frj sum_beer sum_wine  sum_alcot sum_water sum_diet
sum_tea sum_coffee;
by sampleid;
run;

data dir.bevrg_oct;
set merg_bvg_oct;
run;
```

Syntax to make the beverage file compatible with HS file (Gram intake):

```
libname dir "P:\07-SSH-PRC-1331\Nooshin\Aug25\september";
data temp;
set dir.small_merged_sep;
run;
proc contents;run;
proc print;run;
proc freq; tables newcode;run;

data
frdnk sfdnk tswt cofswt milk mlk_hi mlk_lw mlk_oth veg_ju fr_ju beer wine
alc_othr water diet tea coffee;
set temp;

if newcode=1 then output frdnk ;
if newcode=2 then output sfdnk ;
if newcode=3 then output tswt ;
if newcode=4 then output cofswt ;
if newcode=5 then output milk ;
if newcode=6 then output mlk_hi ;
if newcode=7 then output mlk_lw ;
if newcode=8 then output mlk_oth;
if newcode=9 then output veg_ju ;
if newcode=10 then output fr_ju ;
if newcode=11 then output beer ;
if newcode=12 then output wine ;
if newcode=13 then output alc_othr;
if newcode=14 then output water ;
if newcode=15 then output diet ;
if newcode=16 then output tea ;
if newcode=17 then output coffee ;
run;

* Fruits drinks;
proc sort data = frdnk ; by sampleid;
run;
proc means data= frdnk noprint;
by sampleid;
var fidd_wtg;
output out=sum_frdnk sum=fidd_wtg;
run;
data sum_frdnk;
set sum_frdnk;
frdn_wt=fidd_wtg;
run;

data sum_frdnk;
set sum_frdnk;
keep sampleid frdn_wt;
run;
proc sort data=sum_frdnk;
by sampleid;
run;

* soft  drinks;
```

```
proc sort data = sfdnk ; by sampleid;
run;
proc means data= sfdnk noprint;
by sampleid;
var fidd_wtg;
output out=sum_sfdnk sum=fidd_wtg;
run;
data sum_sfdnk;
set sum_sfdnk;
sfdn_wt=fidd_wtg;
run;

data sum_sfdnk;
set sum_sfdnk;
keep sampleid sfdn_wt;
run;
proc sort data=sum_sfdnk;
by sampleid;
run;


* tea-sweetened;
proc sort data = tswt ; by sampleid;
run;
proc means data= tswt noprint;
by sampleid;
var fidd_wtg;
output out=sum_tswt sum=fidd_wtg;
run;
data sum_tswt;
set sum_tswt;
tswt_wt=fidd_wtg;
run;

data sum_tswt;
set sum_tswt;
keep sampleid tswt_wt;
run;
proc sort data=sum_tswt;
by sampleid;
run;


*coffee_sweetened;
proc sort data = cofswt ; by sampleid;
run;
proc means data= cofswt noprint;
by sampleid;
var fidd_wtg;
output out=sum_cfswt sum=fidd_wtg;
run;
data sum_cfswt;
set sum_cfswt;
cfswt_wt=fidd_wtg;
run;
```

```
data sum_cfswt;
set sum_cfswt;
keep sampleid cfswt_wt;
run;
proc sort data=sum_cfswt;
by sampleid;
run;

*plain milk;
proc sort data = milk ; by sampleid;
run;
proc means data= milk noprint;
by sampleid;
var fidd_wtg;
output out=sum_milk sum=fidd_wtg;
run;
data sum_milk;
set sum_milk;
milk_wt=fidd_wtg;
run;

data sum_milk;
set sum_milk;
keep sampleid milk_wt;
run;
proc sort data=sum_milk;
by sampleid;
run;


*milk-based high-fat;
proc sort data = mlk_hi ; by sampleid;
run;
proc means data= mlk_hi noprint;
by sampleid;
var fidd_wtg;
output out=sum_mlkh sum=fidd_wtg;
run;
data sum_mlkh;
set sum_mlkh;
mlkh_wt=fidd_wtg;
run;

data sum_mlkh;
set sum_mlkh;
keep sampleid mlkh_wt;
run;
proc sort data=sum_mlkh;
by sampleid;
run;

*milk_based low_fat;
proc sort data = mlk_lw  ; by sampleid;
run;
proc means data= mlk_lw noprint;
by sampleid;
var fidd_wtg;
```

```
output out=sum_mlkl sum=fidd_wtg;
run;
data sum_mlkl;
set sum_mlkl;
mlkl_wt=fidd_wtg;
run;

data sum_mlkl;
set sum_mlkl;
keep sampleid mlkl_wt;
run;
proc sort data=sum_mlkl;
by sampleid;
run;

*other types of milk;

proc sort data = mlk_oth  ; by sampleid;
run;
proc means data= mlk_oth noprint;
by sampleid;
var fidd_wtg;
output out=sum_mlko sum=fidd_wtg;
run;
data sum_mlko;
set sum_mlko;
mlko_wt=fidd_wtg;
run;

data sum_mlko;
set sum_mlko;
keep sampleid mlko_wt;
run;
proc sort data=sum_mlko;
by sampleid;
run;

*vegatable juice;
proc sort data = veg_ju  ; by sampleid;
run;
proc means data= veg_ju noprint;
by sampleid;
var fidd_wtg;
output out=sum_vegj sum=fidd_wtg;
run;
data sum_vegj;
set sum_vegj;
vegj_wt=fidd_wtg;
run;

data sum_vegj;
set sum_vegj;
keep sampleid vegj_wt;
run;
proc sort data=sum_vegj;
by sampleid;
run;
```

```
*fruit juice;
proc sort data = fr_ju  ; by sampleid;
run;
proc means data= fr_ju noprint;
by sampleid;
var fidd_wtg;
output out=sum_frj sum=fidd_wtg;
run;
data sum_frj;
set sum_frj;
frj_wt=fidd_wtg;
run;

data sum_frj;
set sum_frj;
keep sampleid frj_wt;
run;
proc sort data=sum_frj;
by sampleid;
run;

*beer;
proc sort data = beer  ; by sampleid;
run;
proc means data= beer noprint;
by sampleid;
var fidd_wtg;
output out=sum_beer  sum=fidd_wtg;
run;
data sum_beer ;
set sum_beer ;
beer_wt=fidd_wtg;
run;

data sum_beer;
set sum_beer;
keep sampleid beer_wt;
run;
proc sort data=sum_beer;
by sampleid;
run;

*wine;
proc sort data = wine  ; by sampleid;
run;
proc means data= wine  noprint;
by sampleid;
var fidd_wtg;
output out=sum_wine   sum=fidd_wtg;
run;
data sum_wine  ;
set sum_wine  ;
wine_wt=fidd_wtg;
run;

data sum_wine ;
```

```
set sum_wine ;
keep sampleid wine_wt;
run;
proc sort data=sum_wine ;
by sampleid;
run;

*other alcoholic beverages;
proc sort data = alc_othr  ; by sampleid;
run;
proc means data= alc_othr  noprint;
by sampleid;
var fidd_wtg;
output out=sum_alcot   sum=fidd_wtg;
run;
data sum_alcot  ;
set sum_alcot  ;
alcot_wt=fidd_wtg;
run;

data sum_alcot ;
set sum_alcot ;
keep sampleid alcot_wt;
run;
proc sort data=sum_alcot ;
by sampleid;
run;

*water;
proc sort data = water  ; by sampleid;
run;
proc means data= water   noprint;
by sampleid;
var fidd_wtg;
output out=sum_water    sum=fidd_wtg;
run;
data sum_water   ;
set sum_water   ;
water_wt=fidd_wtg;
run;

data sum_water  ;
set sum_water  ;
keep sampleid water_wt;
run;
proc sort data=sum_water  ;
by sampleid;
run;

*diet drinks;
proc sort data = diet  ; by sampleid;
run;
proc means data= diet   noprint;
by sampleid;
var fidd_wtg;
output out=sum_diet    sum=fidd_wtg;
run;
```

```
data sum_diet   ;
set sum_diet   ;
diet_wt=fidd_wtg;
run;

data sum_diet  ;
set sum_diet  ;
keep sampleid diet_wt;
run;
proc sort data=sum_diet  ;
by sampleid;
run;

*tea;
proc sort data = tea  ; by sampleid;
run;
proc means data= tea   noprint;
by sampleid;
var fidd_wtg;
output out=sum_tea    sum=fidd_wtg;
run;
data sum_tea   ;
set sum_tea   ;
tea_wt=fidd_wtg;
run;

data sum_tea  ;
set sum_tea  ;
keep sampleid tea_wt;
run;
proc sort data=sum_tea  ;
by sampleid;
run;

*coffee;
proc sort data = coffee ; by sampleid;
run;
proc means data= coffee   noprint;
by sampleid;
var fidd_wtg;
output out=sum_coffee    sum=fidd_wtg;
run;
data sum_coffee   ;
set sum_coffee  ;
coffee_wt=fidd_wtg;
run;

data sum_coffee  ;
set sum_coffee  ;
keep sampleid coffee_wt;
run;
proc sort data=sum_coffee ;
by sampleid;
run;


data merg_bvg_sep;
```

```
merge  sum_frdnk sum_sfdnk sum_tswt sum_cfswt sum_milk sum_mlkh sum_mlkl
sum_mlko sum_vegj sum_frj sum_beer sum_wine  sum_alcot sum_water sum_diet
sum_tea sum_coffee;
by sampleid;
run;

data dir.bevrg_sep;
set merg_bvg_sep;
run;
```

Cluster analysis, water included:

```
libname mysas 'P:\07-SSH-PRC-1331\Nooshin\Aug25\september';
libname nooshin 'P:\07-SSH-PRC-1331\Nooshin\Aug25\september';
libname clust 'P:\07-SSH-PRC-1331\Nooshin\Aug25\september';
libname finclust'P:\07-SSH-PRC-1331\Nooshin\Aug25\september';
run;
data nooshin;
set 'P:\07-SSH-PRC-1331\Nooshin\Aug25\september\final_adults';
run;
Data test;
      set nooshin;
if frdn_wt > 2260 then flag_1=1;
if sfdn_wt > 3124 then flag_2=1;
if tswt_wt > 939 then flag_3=1;
if cfswt_wt > 456 then flag_4=1;
if milk_wt > 2179 then flag_5=1;
if mlkh_wt > 1305 then flag_6=1;
if mlkl_wt > 820 then flag_7=1;
if mlko_wt > 328 then flag_8=1;
if vegj_wt > 583 then flag_9=1;
if frj_wt > 1818 then flag_10=1;
if beer_wt > 4003 then flag_11=1;
if wine_wt > 983 then flag_12=1;
if alcot_wt > 885 then flag_13=1;
if water_wt > 7203 then flag_14=1;
if diet_wt > 1226 then flag_15=1;
if tea_wt > 2102 then flag_16=1;
if coffee_wt > 3491 then flag_17=1;
flag_any=.;
if flag_1=1 or flag_2=1 or flag_3=1 or flag_4=1 or flag_5=1 or flag_6=1 or
flag_7=1 or flag_8=1
or flag_9=1 or flag_10=1 or flag_11=1 or flag_12=1 or flag_13=1 or flag_14=1
or flag_15=1 or
flag_16=1 or flag_17=1 then flag_any=1;
run;
proc fastclus data=test maxclusters=40
maxiter=0 cluster=clus40 out=clust.prelim_40_up
outseed=clust.seed_k40_up drift;
where flag_any ne 1;
var frdn_wt sfdn_wt tswt_wt cfswt_wt milk_wt mlkh_wt mlkl_wt mlko_wt vegj_wt
frj_wt beer_wt
wine_wt alcot_wt water_wt diet_wt tea_wt coffee_wt;
run;

title "Prelim cluster adults,exluding>=5SD";
data clusgen;
set clust.prelim_40_up;
/*
proc freq data=clusgen (where=(clus40=11));
table sampleid*   frdn_wt* sfdn_wt* tswt_wt* cfswt_wt* milk_wt* mlkh_wt*
mlkl_wt*
mlko_wt* vegj_wt* frj_wt* beer_wt* wine_wt* alcot_wt* water_wt* diet_wt*
tea_wt*
coffee_wt/list missing;run;
```

```
proc freq data=clusgen (where=(clus40=17));
table sampleid*   frdn_wt* sfdn_wt* tswt_wt* cfswt_wt* milk_wt* mlkh_wt*
mlkl_wt*
mlko_wt* vegj_wt* frj_wt* beer_wt* wine_wt* alcot_wt* water_wt* diet_wt*
tea_wt*
coffee_wt/list missing;run;
proc freq data=clusgen (where=(clus40=20));
table sampleid*   frdn_wt* sfdn_wt* tswt_wt* cfswt_wt* milk_wt* mlkh_wt*
mlkl_wt*
mlko_wt* vegj_wt* frj_wt* beer_wt* wine_wt* alcot_wt* water_wt* diet_wt*
tea_wt*
coffee_wt/list missing;run;
proc freq data=clusgen (where=(clus40=30));
table sampleid*   frdn_wt* sfdn_wt* tswt_wt* cfswt_wt* milk_wt* mlkh_wt*
mlkl_wt*
mlko_wt* vegj_wt* frj_wt* beer_wt* wine_wt* alcot_wt* water_wt* diet_wt*
tea_wt*
coffee_wt/list missing;run;
proc freq data=clusgen (where=(clus40=33));
table sampleid*   frdn_wt* sfdn_wt* tswt_wt* cfswt_wt* milk_wt* mlkh_wt*
mlkl_wt*
mlko_wt* vegj_wt* frj_wt* beer_wt* wine_wt* alcot_wt* water_wt* diet_wt*
tea_wt*
coffee_wt/list missing;run;
proc freq data=clusgen (where=(clus40=39));
table sampleid*   frdn_wt* sfdn_wt* tswt_wt* cfswt_wt* milk_wt* mlkh_wt*
mlkl_wt*
mlko_wt* vegj_wt* frj_wt* beer_wt* wine_wt* alcot_wt* water_wt* diet_wt*
tea_wt*
coffee_wt/list missing;run;
*/
/*beginning n=13584, after flagging adults>5SD from any beverage group*/
if clus40=11 then delete;
if clus40=17 then delete;
if clus40=20 then delete;
if clus40=30 then delete;
if clus40=33 then delete;
if clus40=39 then delete;
run;
/*deletes 15 people, from 6 clusters,n=13569*/
/*****************************************************
Run fastclus on this data set with no outliers, running sets of 3-20 clusters
*****************************************************/

%macro nclus(n);
proc fastclus data=clusgen maxclusters=&n maxiter=50 cluster=clus_40
out=clust.noout_40_&n outseed=clust.seed_40up_&n drift;
var frdn_wt sfdn_wt tswt_wt cfswt_wt milk_wt mlkh_wt mlkl_wt mlko_wt vegj_wt
frj_wt beer_wt
wine_wt alcot_wt water_wt diet_wt tea_wt coffee_wt;
title "No outliers using clusters maxclusters=&n, maxiter=50";
run;
%mend nclus;
%macro doclus (n1,n2);
%do i=&n1 %to &n2;
%nclus (&i);
%end;
```

```
%mend doclus;
%doclus (3,20);
run;

/*PLOTS*/

proc fastclus data=clusgen out=clust maxclusters=6 maxiter=50;
var frdn_wt sfdn_wt tswt_wt cfswt_wt milk_wt mlkh_wt mlkl_wt mlko_wt vegj_wt
frj_wt beer_wt
wine_wt alcot_wt water_wt diet_wt tea_wt coffee_wt;run;

proc candisc data=clust out=can noprint;
class cluster;
var frdn_wt sfdn_wt tswt_wt cfswt_wt milk_wt mlkh_wt mlkl_wt mlko_wt vegj_wt
frj_wt beer_wt
wine_wt alcot_wt water_wt diet_wt tea_wt coffee_wt;
legend1 frame cframe=ligr label=none cborder=black
position=center value=(justify=center);
axis1 label=(angle=90 rotate=0) minor=none;
axis2 minor=none;
symbol1 value=dot
height=.5;
proc gplot data=can;
plot can2*can1=cluster/frame cframe=ligr
legend=legend1 vaxis=axis1 haxis=axis2;run;

/************************
Adding outliers back into cluster set 6
************************/
proc fastclus data=clust.final_adults maxclusters=6 maxiter=0
cluster=clus_6 out=finclust.finalup_6 seed=finclust.seed_40up_6 replace=none;
var frdn_wt sfdn_wt tswt_wt cfswt_wt milk_wt mlkh_wt mlkl_wt mlko_wt vegj_wt
frj_wt beer_wt
wine_wt alcot_wt water_wt diet_wt tea_wt coffee_wt;
title "updated final adults 6 cluster set with replace=none";
run;
```

Cluster analysis, water excluded:

```
libname mysas 'P:\07-SSH-PRC-1331\Nooshin\Aug25\september';
libname nooshin 'P:\07-SSH-PRC-1331\Nooshin\Aug25\september';
libname clust 'P:\07-SSH-PRC-1331\Nooshin\Aug25\september';
libname finclust'P:\07-SSH-PRC-1331\Nooshin\Aug25\september';
run;
data nooshin;
set 'P:\07-SSH-PRC-1331\Nooshin\Aug25\september\final_adults';
run;
Data test;
set nooshin;
if frdn_wt > 2260 then flag_1=1;
if sfdn_wt > 3124 then flag_2=1;
if tswt_wt > 939 then flag_3=1;
if cfswt_wt > 456 then flag_4=1;
if milk_wt > 2179 then flag_5=1;
if mlkh_wt > 1305 then flag_6=1;
if mlkl_wt > 820 then flag_7=1;
if mlko_wt > 328 then flag_8=1;
if vegj_wt > 583 then flag_9=1;
if frj_wt > 1818 then flag_10=1;
if beer_wt > 4003 then flag_11=1;
if wine_wt > 983 then flag_12=1;
if alcot_wt > 885 then flag_13=1;

if diet_wt > 1226 then flag_15=1;
if tea_wt > 2102 then flag_16=1;
if coffee_wt > 3491 then flag_17=1;
flag_any=.;
if flag_1=1 or flag_2=1 or flag_3=1 or flag_4=1 or flag_5=1 or flag_6=1 or
flag_7=1 or flag_8=1
or flag_9=1 or flag_10=1 or flag_11=1 or flag_12=1 or flag_13=1  or flag_15=1
or
flag_16=1 or flag_17=1 then flag_any=1;
run;
proc fastclus data=test maxclusters=40
maxiter=0 cluster=clus40 out=clust.prelim_40_up_nowater
outseed=clust.seed_k40_up_nowater drift;
where flag_any ne 1;
var frdn_wt sfdn_wt tswt_wt cfswt_wt milk_wt mlkh_wt mlkl_wt mlko_wt vegj_wt
frj_wt beer_wt
wine_wt alcot_wt diet_wt tea_wt coffee_wt;
run;

title "Prelim cluster adults no water,exluding>=5SD";
data clusgen_nowater;
set clust.prelim_40_up_nowater;
/*
proc freq data=clusgen_nowater (where=(clus40=11));
table sampleid*   frdn_wt* sfdn_wt* tswt_wt* cfswt_wt* milk_wt* mlkh_wt*
mlkl_wt*
mlko_wt* vegj_wt* frj_wt* beer_wt* wine_wt* alcot_wt* diet_wt* tea_wt*
coffee_wt/list missing;run;
proc freq data=clusgen_nowater (where=(clus40=17));
```

```
table sampleid*   frdn_wt* sfdn_wt* tswt_wt* cfswt_wt* milk_wt* mlkh_wt*
mlkl_wt*
mlko_wt* vegj_wt* frj_wt* beer_wt* wine_wt* alcot_wt* diet_wt* tea_wt*
coffee_wt/list missing;run;
proc freq data=clusgen_nowater (where=(clus40=18));
table sampleid*   frdn_wt* sfdn_wt* tswt_wt* cfswt_wt* milk_wt* mlkh_wt*
mlkl_wt*
mlko_wt* vegj_wt* frj_wt* beer_wt* wine_wt* alcot_wt* diet_wt* tea_wt*
coffee_wt/list missing;run;
proc freq data=clusgen_nowater (where=(clus40=19));
table sampleid*   frdn_wt* sfdn_wt* tswt_wt* cfswt_wt* milk_wt* mlkh_wt*
mlkl_wt*
mlko_wt* vegj_wt* frj_wt* beer_wt* wine_wt* alcot_wt* water_wt* diet_wt*
tea_wt*
coffee_wt/list missing;run;
proc freq data=clusgen_nowater (where=(clus40=30));
table sampleid*   frdn_wt* sfdn_wt* tswt_wt* cfswt_wt* milk_wt* mlkh_wt*
mlkl_wt*
mlko_wt* vegj_wt* frj_wt* beer_wt* wine_wt* alcot_wt* diet_wt* tea_wt*
coffee_wt/list missing;run;
proc freq data=clusgen_nowater (where=(clus40=38));
table sampleid*   frdn_wt* sfdn_wt* tswt_wt* cfswt_wt* milk_wt* mlkh_wt*
mlkl_wt*
mlko_wt* vegj_wt* frj_wt* beer_wt* wine_wt* alcot_wt* diet_wt* tea_wt*
coffee_wt/list missing;run;
*/
/*beginning n=13609, after flagging adults>5SD from any beverage group*/
if clus40=11 then delete;
if clus40=17 then delete;
if clus40=18 then delete;
if clus40=19 then delete;
if clus40=30 then delete;
if clus40=38 then delete;
run;
/*deletes 10 people, from 6 clusters,n=13599*/
/*****************************************************
Run fastclus on this data set with no outliers, running sets of 3-20 clusters
*****************************************************/

%macro nclus(n);
proc fastclus data=clusgen_nowater maxclusters=&n maxiter=50 cluster=clus_40
out=clust.noout_40_nowater_&n outseed=clust.seed_40up_nowater_&n drift;
var frdn_wt sfdn_wt tswt_wt cfswt_wt milk_wt mlkh_wt mlkl_wt mlko_wt vegj_wt
frj_wt beer_wt
wine_wt alcot_wt  diet_wt tea_wt coffee_wt;
title "No outliers using clusters maxclusters=&n, maxiter=50";
run;
%mend nclus;
%macro doclus (n1,n2);
%do i=&n1 %to &n2;
%nclus (&i);
%end;
%mend doclus;
%doclus (3,20);
run;
```

```
/*PLOTS*/

proc fastclus data=clusgen_nowater out=clust_nowater maxclusters=5 maxiter=50;
var frdn_wt sfdn_wt tswt_wt cfswt_wt milk_wt mlkh_wt mlkl_wt mlko_wt vegj_wt
frj_wt beer_wt
wine_wt alcot_wt diet_wt tea_wt coffee_wt;run;

proc candisc data=clust_nowater out=can noprint;
class cluster;
var frdn_wt sfdn_wt tswt_wt cfswt_wt milk_wt mlkh_wt mlkl_wt mlko_wt vegj_wt
frj_wt beer_wt
wine_wt alcot_wt  diet_wt tea_wt coffee_wt;
legend1 frame cframe=ligr label=none cborder=black
position=center value=(justify=center);
axis1 label=(angle=90 rotate=0) minor=none;
axis2 minor=none;
symbol1 value=dot
height=.5;
proc gplot data=can;
plot can2*can1=cluster/frame cframe=ligr
legend=legend1 vaxis=axis1 haxis=axis2;run;

/************************
Adding outliers back into cluster set 5
************************/
proc fastclus data=clust.final_adults maxclusters=5 maxiter=0
cluster=clus_5 out=finclust.finalup_nowater_5
seed=finclust.seed_40up_nowater_5 replace=none;
var frdn_wt sfdn_wt tswt_wt cfswt_wt milk_wt mlkh_wt mlkl_wt mlko_wt vegj_wt
frj_wt beer_wt
wine_wt alcot_wt diet_wt tea_wt coffee_wt;
title "updated final adults 5 cluster set no water with replace=none";
run;
proc fastclus data=clusgen_nowater out=clust_nowater maxclusters=5 maxiter=50;
var frdn_wt sfdn_wt tswt_wt cfswt_wt milk_wt mlkh_wt mlkl_wt mlko_wt vegj_wt
frj_wt beer_wt
wine_wt alcot_wt diet_wt tea_wt coffee_wt;run;

proc candisc data=clust_nowater out=can noprint;
class cluster;
var frdn_wt sfdn_wt tswt_wt cfswt_wt milk_wt mlkh_wt mlkl_wt mlko_wt vegj_wt
frj_wt beer_wt
wine_wt alcot_wt  diet_wt tea_wt coffee_wt;
legend1 frame cframe=ligr label=none cborder=black
position=center value=(justify=center);
axis1 label=(angle=90 rotate=0) minor=none;
axis2 minor=none;
symbol1 value=dot
height=.5;
proc gplot data=can;
plot can2*can1=cluster/frame cframe=ligr
legend=legend1 vaxis=axis1 haxis=axis2;run;

/************************
Adding outliers back into cluster set 5
************************/
proc fastclus data=clust.final_adults maxclusters=5 maxiter=0
```

```
cluster=clus_5 out=finclust.finalup_nowater_5
seed=finclust.seed_40up_nowater_5 replace=none;
var frdn_wt sfdn_wt tswt_wt cfswt_wt milk_wt mlkh_wt mlkl_wt mlko_wt vegj_wt
frj_wt beer_wt
wine_wt alcot_wt diet_wt tea_wt coffee_wt;
title "updated final adults 5 cluster set no water with replace=none";
run;
```