

A study on the modeling for obesity

by

Serena Sungyoung Lim

A thesis submitted to the Faculty of Graduate Studies in partial fulfillment

of the requirements for the Master of Science degree.

Department of

Applied Computer Science

University of Winnipeg

Winnipeg, Manitoba, Canada

December 2021

Abstract

This study primarily aims to develop an Agent Based Model (ABM) that can simulate the obesity rates based on statistical analysis and to find out how obesity is affected by risk factors in a Canadian environment. As obesity can have many causes, it is assumed that various risk factors, not just a decisive one, have an influence on obesity and they interact with one another. Therefore, unlike most previous studies, I approached the obesity problem as a Complex - Adaptive System (CAS).

The data used for this study was provided by Statistics Canada, and the Canadian Community Health Survey (CCHS). This survey is a cross-sectional survey that collects self-reported information related to health status, health care utilization, and health determinants for the Canadian population.

To build the Obesity ABM, it is necessary to find out which risk factors are closely associated with obesity and to what extent they interact with one another. Twelve categories of factors that are expected to influence the obesity rate were chosen on the basis of the related works. Through the statistical data analysis carried out, the main factors and variables for obesity were identified and their respective mathematical relationships obtained. From this, two categories that have several sub-factors for the obesity model were chosen.

I implemented statistical data analysis on the CCHS dataset to see the interrelationship among the factors. Also, I implemented a year-to-year analysis that can show how people change their obesity status each year. Based on the data analysis result, I defined rules for how each risk factor changes each year. These rules are applied to the obesity model using NetLogo.

The architecture of obesity model implementation consists of three main parts: The population module, the risk factor module, and the results module.

Performance evaluation was conducted to examine whether the obesity model can simulate the obesity rate. For this evaluation, the data of CCHS from 2009 to 2014 and the result of the obesity model which is generated by simulation are compared. Model calibration was executed to fit the actual data to the model test result. The result of the model test shows that the percentage error is less than 5%. This means that the obesity model has high validity in predicting obesity for each risk factor.

The obesity ABM is a useful tool to find out the risk factors related to obesity and their relationships in the Canadian population. Thus, this model can potentially assist to improve obesity management at various levels. At the individual level, everyone can find what kinds of strategies are best fit to improve her/his physical condition. Also, at a government or community level, it could help develop policies for people to continue to implement these strategies well. This will lead to reducing the associated social costs and help to promote national health.

Acknowledgments

Through writing this thesis as a computer science graduate student, I believe that I have achieved a lot of personal and intellectual growth. Without the help of the following people, this thesis would not have been possible to complete.

First and foremost, I would like to thank my supervisor Professor Sergio G. Camorlinga for his guidance, encouragement, and support from the beginning to the end. He always provides me with appropriate guidance at each stage of the study. His insightful feedback pushed me to improve my thinking and brought my thesis to a higher level. He kindly advised and encouraged me to continue to study, especially when I was frustrated with a problem during my research. I could not have finished this thesis without him.

My deepest gratitude goes to my thesis committee members. I am deeply grateful for their suggestions and guidance. A big thank you to Professor Yangjun Chen and Professor Carson K. Leung. Their deep and insightful feedback has given me in-depth expertise in my thesis.

I want to express my deep love and gratitude to my parents for their endless support and encouragement. Their unstoppable faith and support are what keeps me moving forward. Their financial support and spiritual encouragement have been the source of my continued studies. I also want to thank my sister for her emotional support and encouragement that makes me feel like I am not away from home.

Finally, I would like to thank all the professors and friends who helped me write this thesis and you who are reading this article.

Contents

Abstract.....	i
Acknowledgments	iii
List of Figures.....	vi
List of Tables	vii
Chapter 1. Introduction	1
1.1 Motivation and problem statement	1
1.2 Research Objectives	2
1.3 Organization of Thesis.....	3
Chapter 2. Related Work	4
2.1 Complex adaptive system	4
2.2. Agent-based model.....	5
2.3 Application of ABM in healthcare.....	6
2.4 Factors of obesity	7
2.5 Statistical analysis methods and related works.....	9
Chapter 3. Conceptual Design	12
3.1 The complexity of obesity.....	12
3.2 Risk Factor for obesity	14
3.3 Obesity as an Agent-Based Model.....	15
Chapter 4. Data analysis.....	18
4.1 Methodology	18
4.2 CCHS Microdata File Structure.....	20
4.3 Hypothesis.....	23
4.4 Statistical Analysis	25
4.4.1 Normal Test	25

4.4.2 MANOVA Test.....	28
4.4.3 Multinomial Logistic Regression Test.....	29
4.5 Mathematical Relationship Amongst Factors	32
4.6 Year-to-year analysis	34
4.7 Obesity risk factor selection.....	36
Chapter 5. Model Implementation	37
5.1 Obesity model development process.....	37
5.2 Architecture of the obesity model implementation.....	39
5.2.1. Population module	40
5.2.2. Risk factor module	40
5.2.3. Result module	40
Chapter 6. Performance Evaluation	42
6.1 Evaluation method	42
6.2 Model calibration and validation	43
6.3 Model Experiment Sets.....	44
6.3.1 Experiment scenario 1: What are the risk factors for increasing obesity? Comparison with each risk independently	44
6.3.2 Experiment scenario 2: What are the risk factors for increasing obesity? Comparison between food consumption and physical activities factors.....	45
6.3.3 Experiment scenario 3: What are the risk factors for increasing obesity? Combination of risk factors (food consumption and physical activities).....	46
Chapter 7. Conclusion and Future Work.....	48
7.1 Summary of the study.....	48
7.2. Strengths of the CAS-based obesity model.....	51
7.3 The limitations of obesity model and future work.....	52
References	54

List of Figures

Figure 3.1. The CAS-based obesity model	12
Figure 3.2. Risk factors for obesity	14
Figure 3.3 the interrelationship among risk factors: obesity, income level, and fruits and vegetable consumption.....	15
Figure 3.4. The structure of the obesity model as an ABM	16
Figure 5.1 Implementation steps for the obesity model.....	37
Figure 5.2 Flow chart for the obesity model	39
Figure 6.1 Dataset split process	43
Figure 6.2 The influence of the risk factor in the obesity model	45

List of Tables

Table 4.1 CCHS microdata file contents	21
Table 4.2 Factors with codes for CCHS microdata file	22
Table 4.3 Normal Test for 2009 to 2014	25
Table 4.4 Multinomial Logistic Regression Test for variables from 2009 to 2014	31
Table 4.5 Main factors for the obesity model	32
Table 4.6 FVCDVEG year to year analysis for 2009-2010.....	35
Table 4.7 FVCDVEG year to year analysis for 2010.....	35
Table 4.8 Risk factors for Obesity model.....	36
Table 6.1 Test result.....	43

Chapter 1.

Introduction

The primary concern of this research is to develop an agent-based model (ABM) to simulate the obesity rate and to see how the individual's (agent's) interaction affects society as a whole (system). The aim of this chapter is to represent the general context of my research. First, I start with the research motivation and problem statement that I have addressed in this research. Then, I present the objective of my research. Lastly, I provide the organization for this thesis.

1.1 Motivation and problem statement

Obesity is a complex disorder that involves an excessive amount of body fat. It is a major public health concern across the world. This is because obesity increases the risk of disease and health problems, including diabetes, high blood pressure, and cardiovascular disease. Over the past decade, the increase in the rate of the obese is accelerating rapidly. It increased as a global epidemic. As reported by the World Health Organization (WHO), the number of obese people has almost tripled since 1975. In 2016, more than 1.9 billion adults, 18 years and older, were overweight. Of these over 650 million were obese. Overall, 39% of adults aged 18 years and over were overweight in 2016, and 13% were obese¹ [WHO21]. Moreover, recent research shows ten times increase in childhood and adolescent obesity for the last 40 years [ABA17]. Overweight and obesity are linked to more deaths worldwide than underweight. In detail, raised

¹ A person with a Body Mass Index (BMI) of less than 18.5 is considered underweight, a person with a Body Mass Index (BMI) of 18.5 to 25.0 is considered normal weight, A person with a Body Mass Index (BMI) of 25.0 to 30 is considered overweight, and a person with a BMI over 30 is considered obese.

Body Mass Index (BMI) is a major risk factor for noncommunicable diseases such as cardiovascular diseases, diabetes, musculoskeletal disorders, and cancers.

Furthermore, obesity places a huge number of economic impacts on public health expenditure. In Canada, on average, health care costs are 25% higher for obese people than normal-weight people. In 2006, adult obesity in Canada accounted for \$3.9 billion in direct health care costs and \$3.2 billion in indirect costs (e.g., costs related to disability and lost productivity due to illness or premature death) [JAN13]. Thus, if we can not stop the spread of obesity, the burden of health expenditure will gradually increase, and this leads to a vicious circle.

To solve this problem, it is very important to find out what are the causes of obesity. Most of the previous studies are focused on finding out one risk factor for obesity. However, obesity is a complex disorder going with an excessive amount of body fat. The factors that influence obesity are not the only one, but various, and the factors interact with each other. In this study, a variety of factors affecting obesity and their inter-relationship are considered by employing statistical analysis from a complex adaptive system perspective. This model can potentially assist to save some associated costs from obesity by helping to prevent obesity and improve obesity management.

1.2 Research Objectives

In this study, I assume that the factors affecting obesity have a complex interrelation with one another. I approach the topic through three objectives. The first objective is to find out the risk factors that affect obesity. For selecting factors, I selected several factors based on related works, and then statistical data was collected. Using the collected data, I conducted the Normal

Test and MANOVA Test to select factors that satisfy normal distribution, linearity, and homogeneity of variances and covariances. The second objective is to understand the complex independencies among risk factors. I ran the Multinomial Logistic Regression Test to find out the relationship between obesity and selected factors. The last objective is to develop an agent-based model that can simulate the obesity rate based on statistical analysis and to see how the individuals (agents) interaction affects society's obesity (system).

1.3 Organization of Thesis

This thesis is organized as follows. Chapter 1 presents the purpose and motivation of the thesis. Chapter 2 provides an overview of related works including complex adaptive systems, obesity, and agent-based modeling. Chapter 3 describes the conceptual design of the obesity model as agent-based modeling. Chapter 4 describes the data analysis. Chapter 5 provides the implementation and simulation of the obesity model. Chapter 6 provides the performance of the obesity model. Chapter 7 includes the conclusion and the suggestions for future works.

Chapter 2.

Related Work

In this chapter, I provide a broad overview of the background material and introduce previous research. A fundamental description of complex adaptive systems, the factors of obesity, the statistical analysis methods, and related works are described next.

2.1 Complex adaptive system

Definition of Complex Adaptive System

According to Lichtenstein [LIC00], complexity researchers have identified four basic assumptions underlying complex systems:

1. Dynamics-Complex systems are dynamic and changing constantly.
2. Not reducible—Due to the entwined nature of the elements, it is insufficient to focus on the effects of the single element as the system as such cannot be reduced to them.
3. Interdependencies—Causality is interdependent, therefore causality in complex systems cannot be explained by a linear model.
4. Non-proportionality—The effect of the input factor is not proportional to the strength of that factor's input. Due to the non-proportionality, small inputs can have large effects, whereas large inputs might hardly change the outcome.

In complex systems research, one of the most powerful tools is the use of computational techniques as they allow a much wider range of exploration of conceptual models. With

computational techniques, I can understand and predict answers to questions arising from my research [MIL09]. CAS can help understand the interrelationships between microscopic processes and macroscopic patterns, and the evolutionary powers that outline the systems [FUC02].

2.2. Agent-based model

Agent-based modeling (ABM) is a simulation approach that consists of an interactive set of agents [WOO09] that represent system components and interactions in a particular environment. ABM offers a theoretical framework from which to explore complex adaptive social systems [MIL09]. The main components of ABM are agents, environments, and time.

1. Agents

Agents are the first component of ABM. An agent can represent an individual (each person). Each agent(individual) has specific attributes (for example, BMI status, daily consumption of total fruits and vegetables, etc.). And each of the agents changes their obesity status based on a set of rules to define their behaviors (for example, to calculate the probability of changing the agent's obese status, IF overweight agent A consume fruits and vegetables, next year agent A will be normal weight, ELSE agent A's obesity status stays same as overweight) and within an environment specified by the researcher [BRU15].

2. Environments

The environments are the space where the agents are live (for example, we can set two environments to see how agents are changing their status over time in different environments). In ABM, manipulating the environment to see how agents(individuals) change their state over time.

3. Time

In ABM, the simulation will go through several time steps (ticks). For example, in the Obesity model, each tick represents one year. Each year agent's obesity status is updated. For example, obese agent A's obesity status can be changed to normal weight at end of the year (one tick).

To build an ABM, it is needed to identify the agents. In this thesis, each agent represents human individuals. Then, it is needed to set what attributes (risk factors) are included in the agents. Agents have internal states (in my study, BMI status: underweight, normal weight, overweight, obesity, etc.), daily consumption of total fruits and vegetables (less than 5 times, 5 times and more, etc.). By selecting the agent's variable, the agent's behavior can be represented as a finite-state automaton: a state transition occurs whenever the agent interacts with another agent. Then, the rules that determine the interactions among the agents are set. The individual agent may experience learning, adaptation, and reproduction. For example, agents can be heterogeneous and different from other agents (for example, BMI status, daily consumption of total fruits and vegetables, etc.). In this way, a computer simulation generates macro-level patterns and trends from the micro-level.

2.3 Application of ABM in healthcare

Agent-based modeling is a well-known approach for visualizing, analyzing, and informing complex dynamic systems in health care research. While ABM can provide insight into population-level health outcomes and interventions, careful consideration of the limitations and challenges of these models is required to understand their full potential [TRA18].

ABM has been effectively used in epidemiologic investigations of contagious disease's spread. Many studies are suggesting diverse model scales, from city scale [EUB04] [PER09] to global scale ABM [PAR11]. In addition, various simulations and interventions tests with ABM can develop strategies for infectious disease control in government agencies such as and the Centers for Disease Control and Prevention (CDC) [LEE13], [COD15].

According to WHO, NCDs, such as heart disease, diabetes, and cancers are mutually responsible for nearly 70% of deaths worldwide. Therefore, ABM has also been efficiently used in noncommunicable diseases (NCDs) control. Unlike contagious diseases, NCDs are long-term diseases, it is more complex to predict their risk factors and their management [AZI16].

2.4 Factors of obesity

Obesity is a noncommunicable disease (NCDs) and one of the most critical public health concerns in the world. Worldwide, 2.8 million people die each year as a result of being overweight. Being overweight and obese would cause adverse metabolic effects on blood pressure, cholesterol, triglycerides, and insulin resistance. The highest prevalence of overweight among infants and young children is another big problem [WHO14].

In this section, I will introduce more detail on the factors of obesity.

- Alcohol

Pure alcohol has an energy content of 7 kcal/g. Among the many problems related to alcohol consumption, the relationship between alcohol consumption and obesity has been extensively studied over the last few years [BEN13] [SAY11] [YEO10]. For example, drinking too much alcohol in the absence of alcohol dependence can increase the risk of obesity [YEO10].

- Chronic conditions

A chronic condition is a health condition or disease that occurs over time or effects last for a long time. Obesity and overweight are strongly associated with several chronic diseases such as diabetes [MOK03]. They are the main causes of the burden of chronic diseases in the population [MUS19] [MOK03] [PAE02].

- Depression

Depression is a medical illness that can negatively affect a person's mood, thoughts, actions, motivation, feelings, and behavior. According to the Centers for Disease Control and Prevention (CDC), the ratio of obese adults with depression was 10% higher than the ratio of obesity in adults without depression. The relationship between obesity and depression is consistent throughout all age groups of women and men aged 60 and over [PRA14].

- Food choices

The high consumption of processed foods, with high caloric density and low nutrients density, combined with lower fiber consumption, and the low expenditure of energy due to the low levels of practicing physical activity could help in the increase of the BMI [MAL16].

- Income

The prevalence of the body mass index (BMI) increases with higher levels of income in countries. The prevalence of overweight in high and middle-income countries was twice higher than in low-income countries. For obesity, both men and women in low-income countries are more than three times the difference from 7% obesity to 24% of upper and upper countries. In high-income countries, women's obesity was much higher than men's. In low and middle-income countries, women's obesity was about twice that of men [WHO19].

- The practice of physical activities

The association between physical activity and obesity is known to be relevant. Physical activity plays an important role in preventing overweight and obesity in childhood and adolescence and reduces the risk of obesity in adulthood. In many western countries, a large percentage of children and adolescents do not meet the recommended physical activity guidelines, and in general, people with physical activity have lower levels of body fat than those with less activity [HIL11].

- Time spent in sedentary behavior

According to Tremblay et al. sedentary behavior is any waking behavior characterized by an energy expenditure of 1.5 metabolic equivalents (METs), while in a sitting, reclining, or lying posture. Common sedentary behaviors are TV watching, computer use (screen time), driving a car, and reading [TRE17]. Reducing the time spent in sedentary behavior and increasing activity is improving weight loss [EPS95][ROB01].

- Habits of smoking

There is a general perception that smoking protects people from gaining weight and makes people gain weight after quitting smoking. Fear of weight gain is often given as a reason to give up quitting smoking. Smoking cessation may be associated with weight gain. Therefore, smoking cessation should be included in weight management support [DAR15][RED08].

2.5 Statistical analysis methods and related works

Statistical analysis provides the relationship among the risk factors of obesity, and it gives insights to set rules for the obesity model. Several examples are described in the next section.

Colapinto et. al. [COL18] explained trends in the frequency of fruit and vegetable consumption in Canada, and the contribution of fruit juice to these trends, and the correlation between the frequency of fruit and vegetable consumption. The data was obtained from the annual Canadian Community Health Survey from 2007 to 2014 and is related to households over 12 years of age. Using weighted frequencies and cross-tabulations, they estimated the average frequency of fruit and vegetable consumption by socio-demographic characteristics and body mass index age-standardized to the Canadian population in 2014. They used multivariate logistic regressions to investigate the correlation with the frequency of fruit and vegetable intake in 2014. This study used the same database from the Canadian Community Health Survey (CCHS) on the NESSTAR data portal (Statistics Canada) and used the Multinomial Logistic Regression to do the statistical analysis. My study also utilizes the same database from CCHS. Thus, I expect the Multinomial Logistic Regression to work well for my data analysis.

Charlton et. al [CHA14] researched the association between body mass index (BMI) and habitual fruit and vegetable consumption in a sample of Australian adults aged 45 and older who were engaged for the "45 and Up" cohort study. Fruit and vegetable intake was assessed using a validated short question, and weight and height were reported by themselves. To assess the association between fruit and vegetable intake and BMI, multinomial logistic regression analysis was used by gender. The data suggest that improving compliance with fruit and vegetable dietary goals may be a dietary strategy to overcome men's overweight, but overweight and obese women already adhere to these goals. The link between adult fruit and vegetable intake and low weight suggests that improving fruit and vegetable intake is important for the overall dietary pattern of this group of people. This study used self-reported baseline questionnaires similar to my database, and Logistic regression was used to assess the relationship between fruit and vegetable

intake and BMI. A multinomial logistic approach was used for the quartile outcomes, and the study also used the Logistic regression to assess the relationship between fruit and vegetable intake and BMI for men and women. A multinomial logistic approach was used for the quartile outcomes. Since my database consists of various multiple questionnaires similar to Charlton's study, thus I decided to use Multinomial Logistic Regression to analyze my data.

Chapter 3.

Conceptual Design

This chapter aims to provide a framework for analyzing the characteristics of the obesity agent-based model as a complex adaptive system. This chapter is divided as follows: 1. The complexity of obesity, 2. Risk factors for obesity, and 3. Obesity as an agent-based model (ABM).

3.1 The complexity of obesity

Obesity is a CAS that includes a wide range of food choices, health conditions, policy, economic status, sedentary behavior, and biological factors and their complex interrelationships. In this study, I approached the obesity problem as CAS. For an agent (individual), obesity is caused by various risk factors such as excessive energy intake, health status, and physical activities.

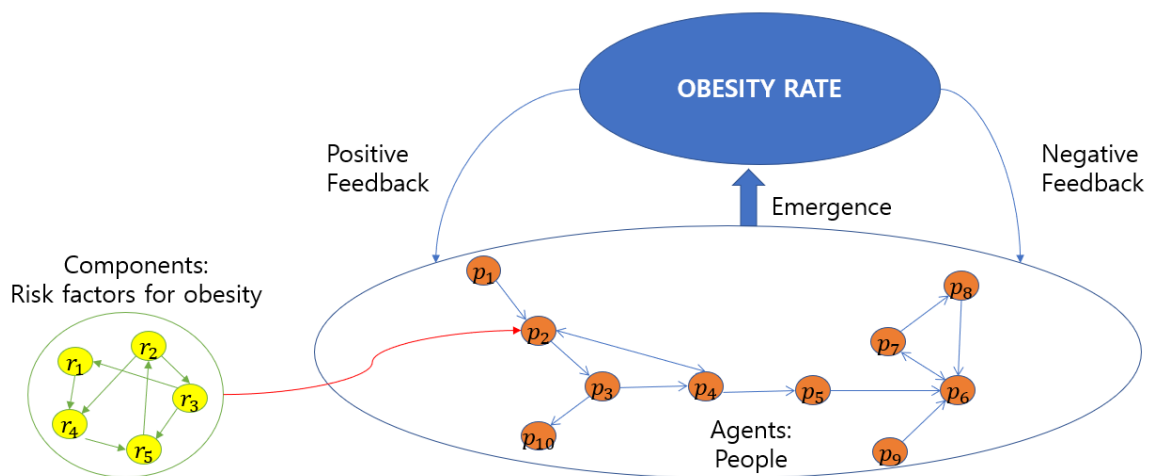


Figure 3.1 The CAS-based Obesity Model

Figure 3.1. shows the proposed CAS-based obesity model in this study. The complex interrelationship among each of the risk factors which give rise to an individual's obesity (yellow circles r_1, r_2 , etc.) is shown in the smaller circle. Then, the individual beings (orange circles p_1, p_2 , etc.) co-exist to emerge with the obesity rates at the community level. The environment is also an important factor in the emergence of obesity at the individual and community levels. For instance, larger entities (e.g. government healthcare organizations, industry, etc.) can have the power to influence the environment either positively or negatively via policies and marketing (e.g. nutrition campaigns, junk food marketing, etc.). In this research, I focus on the obesity risk factor interrelationships and interdependencies with the use of statistical analysis. My final aim is to use the risk factors studied (yellow circles r_1, r_2 , etc.) to build an obesity model based on CAS principles that will help us to simulate obesity rates for adults in a community.

3.2 Risk Factor for obesity

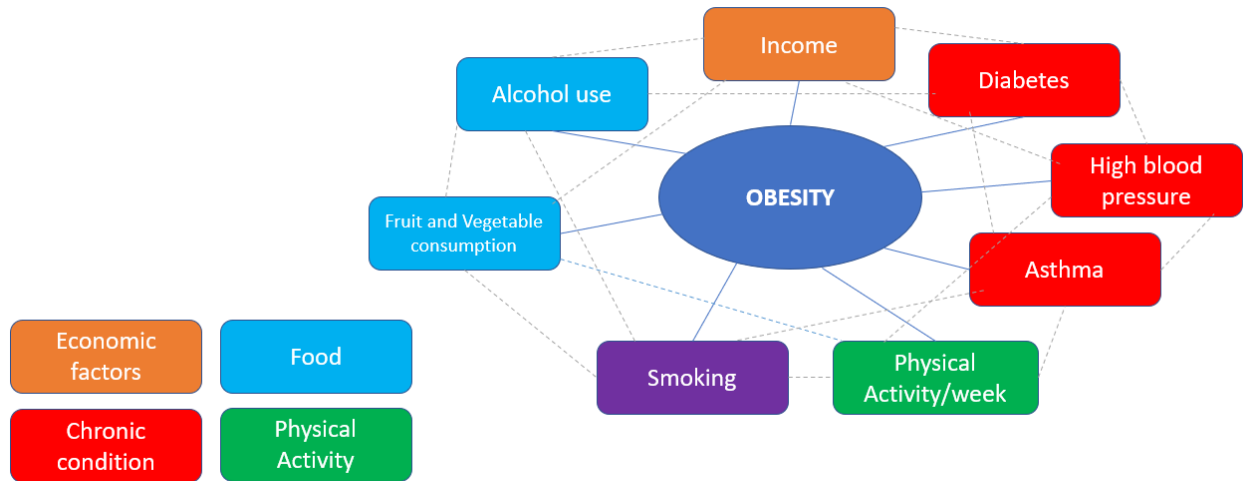


Figure 3.2. Risk factors for Obesity

Risk factors for obesity are selected from those given in chapter 2. Each risk factor's status is changeable for each year. For example, underweight person A's physical activity/week in 2009-2010 is 1(active) but in 2010 it can be changed to 2(inactive). This can affect underweight person A's BMI status. Also, each risk factor can affect the other's status. As described in figure 3.2., each factor is not only related to obesity independently but also affects each other. For example, the patient who has high blood pressure are more likely to have asthma, and a patient who has asthma are also likely to have high blood pressure [ROE10] [CHR16].

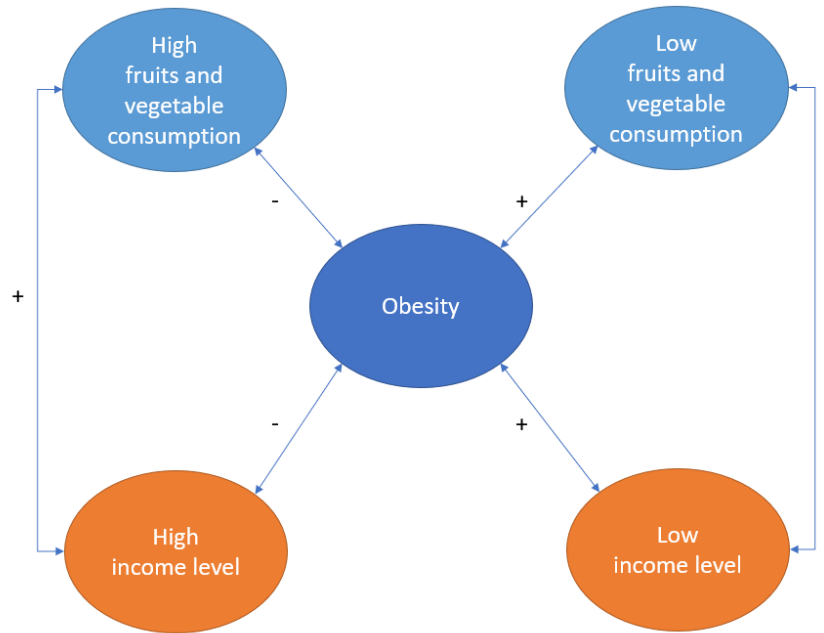


Figure 3.3 the interrelationship among risk factors: obesity, income level, and fruits and vegetable consumption

Figure 3.3 shows the interrelationship among risk factors (obesity, income level, and fruits and vegetable consumption). For example, economic factors such as income level affect food choices. In detail, the low-income level is a major obstacle to increasing fruits and vegetables in diet [HOU15]. This means the low-income level causes low consumption of fruits and vegetables, and this increases the probability of the agent's obese status to be overweight or obese.

3.3 Obesity as an Agent-Based Model

Agent-based modeling (ABM) is a powerful simulation modeling method that can apply to real-world problems. In ABM, a system is modeled as a collection of autonomous decision-making entities called agents. This study aims to build an obesity ABM model and find out how obesity is affected by risk factors (such as fruit and vegetable consumption and physical

activities). In this model, each agent represents human individuals and interacts with the other agents. Each agent evaluates the situation individually and decides based on a set of rules [BON02] [HEN16].

The obesity model consists of three key elements: agent, risk factors, and time. In detail, in this obesity model, an agent can represent each individual in the population and each person's characteristics are set as risk factors (attributes of agents) and set rules for changing their status over time to see the dynamics of obesity.

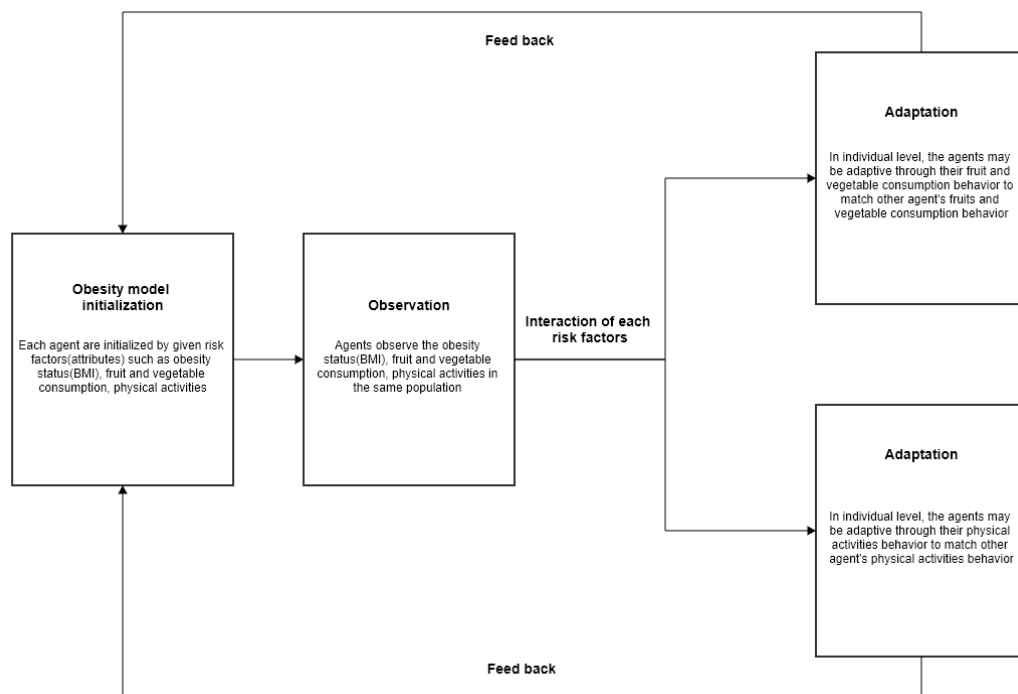


Figure 3.4. The structure of the obesity model as an ABM

Figure 3.4 shows the structure of the obesity model as an agent-based model. First, the obesity model is initialized. To be specific, each individual (agent) is initialized with selected risk factors such as obesity status, fruits and vegetable consumption, and physical activities.

Then each agent is observed their obesity status, fruits and vegetable consumption, and physical activities in the given population. Next, each agent interacts with the risk factors. For example, at the individual level, an agent's physical activity status may change or remain the same (physically active or physically inactive). This adaptation affects the agent's obesity status (feedback). As time goes by, the obesity model will allow us to see how this feedback affects the agent's obesity status. For example, it will show how feedback affects the risk factors and agents, how time evolves in the obesity model, unexpected behaviors that emerge, and more.

Chapter 4.

Data Analysis

In this chapter, firstly, I explain the methodology of my work. Secondly, I conducted the Normal Test, MANOVA Test, and Multinomial Logistic Regression Test for analyzing the data. The data analysis will help to obtain the relevant variables that influence the obesity model including the mathematical relationship among the risk factors, and the risk factor selection to be used for the obesity model. The data used for this study was provided by Statistics Canada, and the Canadian Community Health Survey (CCHS), available on the NESSTAR data portal (Statistics Canada). This survey is a cross-sectional survey that collects self-reported information related to health status, health care utilization, and health determinants for the Canadian population. It relies upon a large sample of respondents and is designed to provide reliable health estimates at the regional level [STA18].

4.1 Methodology

Firstly, a bibliographic search identified human studies of risk factors associated with overweight and obesity [CEN18] [MAY19] [USD16] [WHO18] [WHO21]. Twelve categories of factors that are expected to change the obesity rate are chosen on the basis of the related works in Chapter 2. The twelve categories of factors included were: alcohol use, chronic conditions, changes made to improve health, depression, food choices, fruit and vegetable consumption, general health, income, mood, physical activities, sedentary activities, and smoking. These findings allowed us to identify relevant obesity risk factors in the Canadian Community Health Survey.

It was noticed that the survey questions before 2009 were quite different from those in the period 2009 to 2014. Therefore, I decided to focus on the data from 2009 to 2014 as the survey contained the same factors, variables, questions, and types of answers. Participants with incomplete and missing responses such as “not applicable”, “don’t know”, “refusal” and “not stated” were excluded, i.e., the factors not important for this study were excluded (for example “Contacts with health professionals”), and the remaining list are the factors considered in this research and include:

- Fruit and vegetable consumption

- FVCDVEG - Daily consumption - Other vegetables

- FVCDTOT - Daily consumption - Total fruits and vegetables

- Physical activities

- PAC_7 - Walked to work or school - Past 3 months

- PAC_8 - Bicycled to work or school - Past 3 months

- PACDEE - Daily energy expenditure - Leisure physical activities

- PACFLEI - Participant in leisure physical activity

- PACDFR - Frequency - All leisure physical activities lasting more than 15 minutes

- PACFD - Participant in daily leisure physical activity lasting more than 15 minutes

- PACDPAI - Leisure physical activity index

- PACDLTI - Leisure and transportation physical activity index

- PACDTLE - Daily energy expenditure - Transportation and leisure physical activity

- PACFLTI - Participant in leisure or transportation physical activity

The statistical analysis aimed to identify how the various risk factors contribute to the emergence of an individual's obesity. Ideally, I wanted to obtain a linear mathematical formula that includes the relevant risk factors and their coefficients as shown in Equation 1.

$$\text{Individual obesity} = a r_1 + b r_2 + c r_3 + \dots + n r_n, \quad (\text{Equation 1})$$

where a, b, c, \dots, n are coefficients that represent the risk factor relevance $r_1, r_2, r_3, \dots, r_n$ of the respective obesity risk factor. Risk factors included were fruit and vegetable consumption, and physical activities.

As the main idea was to try to create a mathematical formula to use in the construction of the modeling, it was verified that the best statistical analyses, for the database of those years and subsequent use for the creation of the software, would be to perform the Normal Test, MANOVA and Multivariate Logistic Regression, as described in more details in Section 4.4. To carry out these tests, the statistical analysis program IBM SPSS Statistics, version 19 was used. From this, some statistical tests were performed (Normal Test, Independent samples t-test, and Simple Linear Regression).

4.2 CCHS Microdata File Structure

Canadian Community Health Survey (CCHS) provides a microdata file that contains each person's responses to the survey. The CCHS microdata file is open to researchers from various fields to conduct research to improve health. Table 4.1 shows an example of the microdata file. Each column represents a survey question, and each row represents a participant's answer for a survey item. The numbers represent the answer to a survey item chosen by the participant. For example, in Table 1 the column HWTGISW means Body Mass Index (BMI) classification by a survey participant (underweight, normal weight, overweight, and obese) the first individual

answered ‘2’ to this question. For example, this survey participant answered: I am “normal weight”.

Table 4.1. CCHS microdata file contents

HWTGISW	FVCDJUI	FVCDFRU	FVCDLVEG	FVCDPOT	FVCDCAR	FVCDVEG	...	FVCDTOT
2	1	1	1	1	1	1	...	2
3	1	1	1	1	1	1	...	1
2	1	1	1	1	1	1	...	1
4	1	1	1	1	1	1	...	1
...
2	1	1	1	1	1	1	...	1

In addition, I obtained data from various health-related risk factors such as fruit and vegetable consumption, and physical activities. Also, each risk factor has sub-factors; e.g., the risk factor for physical activities has 10 sub-variables such as participant in leisure physical activity, bicycled to work or school, etc. I only want to include risk factors truly associated with obesity that had been identified by statistical data analyses as described in the next section.

Table 4.2 Factors with codes for CCHS microdata file

Factors	Variable	Codes	Questions	Answers			
HWT: Height and weight	Body mass index (BMI) classification (respondents aged 18 and over) - Self-reported	HWTGISW	BMI class. (18+) / self-report - Intern. standard	1	2	3	4
				Underweight	Normal weight	Overweight	Obese
ALC: Alcohol use	Frequency of drinking alcohol	ALC_2	During the past 12 months, how often did you drink alcoholic beverages?	1	2	3	4
				Less than once a month	Once a month	2 to 3 times a month	Once a week
FVC: Fruit and vegetable consumption	Daily consumption - Fruit juice	FVCDJUI	This variable indicates the frequency per day of consumption of drinks fruit juice.	1	2		
				Less than 5 times	5 times or more		
	Daily consumption - Fruit	FVCDFRU	This variable indicates the frequency per day of consumption of fruit, excluding fruit juices.	1	2		
				Less than 5 times	5 times or more		
	Daily consumption - Green salad	FVCDLAL	This variable indicates the frequency per day of consumption of green salad.	1	2		
				Less than 5 times	5 times or more		
	Daily consumption - Potatoes	FVCDPOT	This variable indicates the frequency per day of consumption of potatoes, excluding French fries, fried potatoes, or potato chips.	1	2		
				Less than 5 times	5 times or more		
	Daily consumption - Carrots	FVCDLAL	This variable indicates the frequency per day of consumption of carrots.	1	2		
			Less than 5 times	5 times or more			
Daily consumption - Other vegetables	FVCDVEG	This variable indicates the respondent's usual daily consumption of other vegetables, excluding carrots, potatoes, or salad.	1	2			
			Less than 5 times	5 times or more			
Daily consumption - Total fruits and vegetables	FVCDTOT	This variable indicates the frequency per day of consumption of fruits and vegetables.	1	2			
			Less than 5 times	5 times or more			
Daily consumption - Total fruits and vegetables	FVCGTOT	This variable classifies the respondent based on the total number of times per day he/she eats fruits and vegetables.	1	2	3		
			Eats fruits and vegetables less than 5 times per day	Eats fruits and vegetables between 5 and 10 times per day	Eats fruits and vegetables more than 10 times per day		

PAC: Physical activities	Walked to work or school - Past 3 months	PAC_7	Time in the past 3 months when you walked to and from work or school?	1 Yes	2 No	
	Bicycled to work or school - Past 3 months	PAC_8	Time in the past 3 months when you bicycled to and from work or school?	1 Yes	2 No	
	Daily energy expenditure - Leisure physical activities	PACDEE	Daily frequency in leisure physical activities	1 Less than 5 times	2 5 times or more	
	Participant in leisure physical activity	PACFLEI	Participant in leisure physical activity	1 Yes	2 No	
	Frequency - All leisure physical activity lasting more than 15 minutes	PACDFR	Frequency of all leisure physical activity > 15 min.	1 Regular	2 Occasional	3 Infrequent
	Participant in daily leisure physical activity lasting more than 15 minutes	PACFD	Participant in daily leisure physical activity > 15 min.	1 Yes	2 No	
	Leisure physical activity index	PACDPAI	Leisure physical activity index	1 Active	2 Moderate active	3 Inactive
	Leisure and transportation physical activity index	PACDLTI	Leisure and transportation physical activity index	1 Active	2 Moderate active	3 Inactive
	Daily energy expenditure - Transportation and leisure physical activity	PACDTLE	Daily frequency in transportation and leisure physical activity	1 Less than 5 times	2 5 times or more	
	Participant in leisure or transportation physical activity	PACFLTI	Participant in leisure or transportation physical activity	1 Yes	2 No	

4.3 Hypothesis

I searched for adequate statistical analyzes of the CCHS data. I found that *Independent T-test* and *Simple Linear Regression* cannot be used as both tests only apply to up to two independent variables (factors), and this does not apply to the CCHS data. The CCHS database contains various independent categorical variables (factors).

Related works reported on Chapter 2 used Multivariate Logistic Regression Analysis --- Colapinto et. al. [COL18] explored the relationship between BMI and fruit and vegetable intake

in the Canadian context, and Charlton et. al. [CHA14], in the Australian context. These studies differ from my study which is comprehensive as it includes a large range of different risk factors.

My study is also applying correlations amongst the risk factors into a CAS-based obesity model.

My goal is to create a mathematical equation for the construction of a model to determine the development of obesity at an individual level. Based on the available datasets, good statistical

analyses would be Normal Test, MANOVA (Multivariate Analysis of Variance), and

Multivariate Logistic Regression Analysis.

4.4 Statistical Analysis

4.4.1 Normal Test

The Normal Test determines if the data distribution presents normality or not. To carry out this test, the statistical analysis program IBM SPSS Statistics was used. As shown in Table 4.3, between 2009 and 2014 all risk factor variables were normally distributed ($p\text{-value} \leq 0.05$). Thus, these variables can be used to perform a MANOVA test.

Table 4.3 Normal Test for 2009 to 2014

Variable	BMI	2009-2010		2010		2011-2012		2012		2013-2014		2014	
		n	p-value	n	p-value	n	p-value	n	p-value	n	p-value	n	p-value
FVCDJUI	U*	297	.000	779	.000	1573	.000	767	.000	1613	.000	753	.000
	N*	7167	.000	17197	.000	34218	.000	16684	.000	33221	.000	16325	.000
	OW*	5838	.000	14129	.000	28151	.000	13890	.000	28845	.000	14206	.000
	OB*	3275	.000	7935	.000	16513	.000	8265	.000	17665	.000	8918	.000
FVCDFRU	U	297	.000	779	.000	1573	.000	767	.000	1613	.000	753	.000
	N	7167	.000	17197	.000	34218	.000	16684	.000	33221	.000	16325	.000
	OW	5838	.000	14129	.000	28151	.000	13890	.000	28845	.000	14206	.000
	OB	3275	.000	7935	.000	16513	.000	8265	.000	17665	.000	8918	.000
FVCDLAL	U	297	.000	779	.000	1573	.000	767	.000	1613	.000	753	.000
	N	7167	.000	17197	.000	34218	.000	16684	.000	33221	.000	16325	.000
	OW	5838	.000	14129	.000	28151	.000	13890	.000	28845	.000	14206	.000
	OB	3275	.000	7935	.000	16513	.000	8265	.000	17665	.000	8918	.000
FVCDPOT	U	297	.000	779	.000	1573	.000	767	.000	1613	.000	753	.000
	N	7167	.000	17197	.000	34218	.000	16684	.000	33221	.000	16325	.000
	OW	5838	.000	14129	.000	28151	.000	13890	.000	28845	.000	14206	.000
	OB	3275	.000	7935	.000	16513	.000	8265	.000	17665	.000	8918	.000
FVCDLAL	U	297	.000	779	.000	1573	.000	767	.000	1613	.000	753	.000
	N	7167	.000	17197	.000	34218	.000	16684	.000	33221	.000	16325	.000
	OW	5838	.000	14129	.000	28151	.000	13890	.000	28845	.000	14206	.000
	OB	3275	.000	7935	.000	16513	.000	8265	.000	17665	.000	8918	.000
FVCDVEG	U	297	.000	779	.000	1573	.000	767	.000	1613	.000	753	.000
	N	7167	.000	17197	.000	34218	.000	16684	.000	33221	.000	16325	.000
	OW	5838	.000	14129	.000	28151	.000	13890	.000	28845	.000	14206	.000
	OB	3275	.000	7935	.000	16513	.000	8265	.000	17665	.000	8918	.000
FVCDTOT	U	297	.000	779	.000	1573	.000	767	.000	1613	.000	753	.000
	N	7167	.000	17197	.000	34218	.000	16684	.000	33221	.000	16325	.000
	OW	5838	.000	14129	.000	28151	.000	13890	.000	28845	.000	14206	.000
	OB	3275	.000	7935	.000	16513	.000	8265	.000	17665	.000	8918	.000
FVCGTOT	U	297	.000	779	.000	1573	.000	767	.000	1613	.000	753	.000
	N	7167	.000	17197	.000	34218	.000	16684	.000	33221	.000	16325	.000
	OW	5838	.000	14129	.000	28151	.000	13890	.000	28845	.000	14206	.000

	OB	3275	.000	7935	.000	16513	.000	8265	.000	17665	.000	8918	.000
PAC_7	U	297	.000	779	.000	1573	.000	767	.000	1613	.000	753	.000
	N	7167	.000	17197	.000	34218	.000	16684	.000	33221	.000	16325	.000
	OW	5838	.000	14129	.000	28151	.000	13890	.000	28845	.000	14206	.000
	OB	3275	.000	7935	.000	16513	.000	8265	.000	17665	.000	8918	.000
PAC_8	U	297	.000	779	.000	1573	.000	767	.000	1613	.000	753	.000
	N	7167	.000	17197	.000	34218	.000	16684	.000	33221	.000	16325	.000
	OW	5838	.000	14129	.000	28151	.000	13890	.000	28845	.000	14206	.000
	OB	3275	.000	7935	.000	16513	.000	8265	.000	17665	.000	8918	.000
PACDEE	U	297	.000	779	.000	1573	.000	767	.000	1613	.000	753	.000
	N	7167	.000	17197	.000	34218	.000	16684	.000	33221	.000	16325	.000
	OW	5838	.000	14129	.000	28151	.000	13890	.000	28845	.000	14206	.000
	OB	3275	.000	7935	.000	16513	.000	8265	.000	17665	.000	8918	.000
PACFLEI	U	297	.000	779	.000	1573	.000	767	.000	1613	.000	753	.000
	N	7167	.000	17197	.000	34218	.000	16684	.000	33221	.000	16325	.000
	OW	5838	.000	14129	.000	28151	.000	13890	.000	28845	.000	14206	.000
	OB	3275	.000	7935	.000	16513	.000	8265	.000	17665	.000	8918	.000
PACDFR	U	297	.000	779	.000	1573	.000	767	.000	1613	.000	753	.000
	N	7167	.000	17197	.000	34218	.000	16684	.000	33221	.000	16325	.000
	OW	5838	.000	14129	.000	28151	.000	13890	.000	28845	.000	14206	.000
	OB	3275	.000	7935	.000	16513	.000	8265	.000	17665	.000	8918	.000
PACFD	U	297	.000	779	.000	1573	.000	767	.000	1613	.000	753	.000
	N	7167	.000	17197	.000	34218	.000	16684	.000	33221	.000	16325	.000
	OW	5838	.000	14129	.000	28151	.000	13890	.000	28845	.000	14206	.000
	OB	3275	.000	7935	.000	16513	.000	8265	.000	17665	.000	8918	.000
PACDPAI	U	297	.000	779	.000	1573	.000	767	.000	1613	.000	753	.000
	N	7167	.000	17197	.000	34218	.000	16684	.000	33221	.000	16325	.000
	OW	5838	.000	14129	.000	28151	.000	13890	.000	28845	.000	14206	.000
	OB	3275	.000	7935	.000	16513	.000	8265	.000	17665	.000	8918	.000
PACDLTI	U	297	.000	779	.000	1573	.000	767	.000	1613	.000	753	.000
	N	7167	.000	17197	.000	34218	.000	16684	.000	33221	.000	16325	.000
	OW	5838	.000	14129	.000	28151	.000	13890	.000	28845	.000	14206	.000
	OB	3275	.000	7935	.000	16513	.000	8265	.000	17665	.000	8918	.000
PACDTLE	U	297	.000	779	.000	1573	.000	767	.000	1613	.000	753	.000
	N	7167	.000	17197	.000	34218	.000	16684	.000	33221	.000	16325	.000
	OW	5838	.000	14129	.000	28151	.000	13890	.000	28845	.000	14206	.000
	OB	3275	.000	7935	.000	16513	.000	8265	.000	17665	.000	8918	.000
PACFLTI	U	297	.000	779	.000	1573	.000	767	.000	1613	.000	753	.000
	N	7167	.000	17197	.000	34218	.000	16684	.000	33221	.000	16325	.000
	OW	5838	.000	14129	.000	28151	.000	13890	.000	28845	.000	14206	.000
	OB	3275	.000	7935	.000	16513	.000	8265	.000	17665	.000	8918	.000
ALC_2	U	297	.000	779	.000	1573	.000	767	.000	1613	.000	753	.000
	N	7167	.000	17197	.000	34218	.000	16684	.000	33221	.000	16325	.000
	OW	5838	.000	14129	.000	28151	.000	13890	.000	28845	.000	14206	.000
	OB	3275	.000	7935	.000	16513	.000	8265	.000	17665	.000	8918	.000

CI: 95% Confidence Interval , *U: Underweight, N: Normal weight, OW: Overweight, OB: Obese

The Independent sample t-test was made to test the null hypothesis that the means of two populations are the same ($H_0: \mu_1 = \mu_2$) when a sample of observations from each population is

available. The observations made on the sample members must all be independent of each other. Another hypothesis (i.e., hypothesis one) means that the two populations are not the same ($H_1: \mu_1 \neq \mu_2$) and the results present the statistical difference between means of two groups [SAB10]. For this test, the "Normal Weight" group means were compared with the means of the "Overweight and Obesity" group. Since the T Student Test is a parametric test, it had to meet some prerequisites [SAB10] as follows:

- $n \geq 30$: the sample for this test had $n=41474$ (n are the sample values of the response and exploratory variables);
- Normal distribution: the sample had a normal distribution ($p\text{-value} \leq 0.05$ in the Tests of Normality).

However, the data have more than two independent variables (group). Thus, it is not possible to perform the Independent Sample t-test. The Simple Linear Regression was made thinking of formulating an equation to estimate the conditional (expected value) of one variable (y), given the values of some other variables (x) from the correlation values between the factors.

This type of test has other prerequisites, but the data have more than two independent variables. Thus, the Simple Linear Regression cannot be executed [SAB10]. Therefore, I reviewed previous works for types of statistical analyses that could be performed for samples with more than two variables to analyze at the same time and if it is possible to see the correlations in the variables. I found other research that supports the statistical analysis that we needed, such as the "Trends and correlates of frequency of fruit and vegetable consumption, 2007 to 2014" [COL18] that used the same database from the Canadian Community Health Survey (CCHS) on the NESSTAR data portal. This work used the Multivariate Logistic

Regression to do a similar statistical analysis. Another research that used something similar statistical analysis was “Fruit and Vegetable Intake and Body Mass Index in a Large Sample of Middle-Aged Australian Men and Women” [CHA14].

4.4.2 MANOVA Test

MANOVA (Multivariate Analysis of Variance) is an extension of *ANOVA* (Analysis of Variance). *MANOVA* tests whether there is any difference in two or more independent vectors of means on a dependent factor—in our study we have one dependent factor (body weight) and twenty independent vectors. To conduct *MANOVA*, three assumptions must be met [FRE08]:

- **Normal Distribution:** The dependent variable should be normally distributed within groups---the Normal test must precede the *MANOVA* test.
- **Linearity:** *MANOVA* assumes that there are linear relationships among all pairs of dependent variables. So, when the relationship deviates from linearity, the power of the analysis will be compromised.
- **Homogeneity of variances and covariances:** In multivariate analysis, since there are multiple dependent variables, it is also required that their intercorrelations (covariances) are homogeneous across the variables of the study.

MANOVA test was performed using SPSS (IBM SPSS Statistics). I found that six variables, daily consumption of fruit juice (FVCDJUI), daily consumption of fruit (FVCDFRU), daily consumption of green salad (FVCDL), daily consumption of potatoes (FVCDPOT), daily consumption of carrots (FVCDCAR), and frequency of drinking alcohol (ALC_2) showed low or zero variance with the dependent variable “Body mass index (BMI) classification” (p-

value ≥ 0.05) and thus were removed. I now can execute the *Multinomial Logistic Regression Test* on the remaining variables.

4.4.3 Multinomial Logistic Regression Test

A regression test explores the relationship between a dependent variable and a set of independent or predictor variables [BEW03]. In our study “normal weight” was set as the reference category to compare with the other groups (underweight, overweight, and obesity). At a p-value ≥ 0.05 , some variables proved to not be significantly associated with underweight or overweight/obesity. Table 4.4 shows the results of the *Multinomial Logistic Regression Test* from the CCHS dataset from 2009 to 2014. The B and Sig. columns represent the logistic constant and the significance value respectively. For example, in 2014, the logistic constant of "obese" for the variable leisure and transportation physical activity index (PACDLTI) is 0.497 and statistically significant since its p-value ≤ 0.05 . This means that the risk of obesity increases as the value (1 = active, 3 = inactive) of the leisure and transportation physical activity index increases. Thus, an inactive person is more likely to be obese than an active person. Significant independent predictors for underweight, overweight and obese status in 2014 are:

- Underweight—walked to work or school for past 3 months (PAC_7), bicycled to work or school for past 3 months (PAC_8), and frequency of all leisure physical activity lasting more than 15 minutes (PACDFR)

- Overweight— daily consumption of total fruits and vegetables (FVCGTOT), walked to work or school for past 3 months (PAC_7), frequency of all leisure physical activity over 15 minutes (PACDFR), leisure and transportation physical activity index (PACDLTI), and daily frequency in transportation and leisure physical activity (PACDTLE)

- Obesity— daily consumption of total fruits and vegetables (FVCGTOT), walked to work or school for past 3 months (PAC_7), frequency of all leisure physical activity over 15 minutes (PACDFR), leisure physical activity index (PACDPAI), leisure and transportation physical activity index (PACDLTI), and daily frequency in transportation and leisure physical activity (PACDTLE)

Table 4.4 Multinomial Logistic Regression Test for variables from 2009 to 2014

BMI	Variables	2009-2010		2010		2011-2012		2012		2013-2014		2014	
		B	Sig.	B	Sig.	B	Sig.	B	Sig.	B	Sig.	B	Sig.
Under weight	Intercept	-4.025	.000	-4.012	.000	-3.404	.000	-3.815	.000	-4.215	.000	-4.205	.000
	FVCDVEG	-.185	.672	-.241	.350	-.394	.033	-.336	.190	.119	.456	.034	.880
	FVCDTOT	-.266	.518	-.225	.296	-.426	.008	-.369	.102	-.105	.537	-.412	.074
	FVCGTOT	.289	.440	.305	.101	.380	.006	.350	.075	.042	.782	.371	.067
	PAC_7	-.191	.255	-.561	.000	-.391	.000	-.361	.001	-.424	.000	-.516	.000
	PAC_8	.358	.083	.698	.000	.449	.000	.551	.000	.565	.000	.713	.000
	PACDEE	-.324	.606	-.313	.397	-.123	.634	.282	.518	.138	.608	.207	.612
	PACFLEI	1.146	.033	-.235	.546	.299	.210	.623	.048	-.175	.535	-.145	.726
	PACDFR	.218	.057	.283	.000	.183	.000	.181	.010	.229	.000	.188	.007
	PACFD	-.051	.793	-.086	.488	-.146	.089	.025	.842	-.132	.123	-.088	.482
	PACDPAI	.125	.683	-.149	.459	.221	.076	.290	.105	.356	.004	.043	.829
	PACDLTI	-.051	.864	.266	.174	-.038	.754	-.189	.272	-.150	.205	.146	.445
	PACDTLE	.105	.863	.181	.613	.135	.592	-.249	.561	-.007	.978	-.151	.706
PACFLTI	-.753	.180	.464	.247	-.166	.511	-.588	.083	.220	.457	.152	.724	
Over weight	Intercept	.285	.194	-.352	.011	-.067	.494	-.160	.250	-.097	.316	-.135	.329
	FVCDVEG	-.269	.048	-.100	.236	-.231	.000	-.216	.008	-.165	.003	-.109	.166
	FVCDTOT	-.300	.023	.039	.614	-.099	.075	.016	.838	-.063	.263	-.038	.639
	FVCGTOT	.093	.444	-.213	.002	-.085	.091	-.205	.005	-.115	.026	-.147	.049
	PAC_7	.178	.001	.235	.000	.205	.000	.238	.000	.221	.000	.215	.000
	PAC_8	-.134	.029	-.063	.132	-.061	.044	-.070	.103	-.058	.061	-.056	.204
	PACDEE	.286	.194	.188	.203	.195	.052	.312	.030	.219	.029	.254	.077
	PACFLEI	.145	.643	-.109	.531	-.181	.149	-.239	.200	-.113	.377	-.184	.318
	PACDFR	-.042	.246	-.018	.419	-.030	.066	-.034	.147	-.042	.010	-.099	.000
	PACFD	.042	.444	-.018	.626	-.031	.233	-.053	.144	.016	.530	.043	.244
	PACDPAI	-.216	.046	-.101	.154	-.087	.077	.036	.613	-.058	.241	-.085	.228
	PACDLTI	.307	.004	.208	.003	.192	.000	.083	.229	.173	.000	.212	.002
	PACDTLE	-.417	.055	-.249	.087	-.328	.001	-.373	.009	-.348	.000	-.342	.016
PACFLTI	-.193	.547	.147	.410	.266	.039	.299	.119	.085	.520	.124	.516	
Obese	Intercept	-.836	.002	-1.297	.000	-.924	.000	-.981	.000	-.960	.000	-.729	.000
	FVCDVEG	-.211	.222	-.031	.771	-.188	.009	-.138	.161	-.043	.531	-.069	.484
	FVCDTOT	-.477	.004	-.078	.417	-.239	.001	-.190	.049	-.053	.456	-.020	.846
	FVCGTOT	.193	.204	-.154	.077	.004	.949	-.042	.631	-.187	.005	-.253	.009
	PAC_7	.098	.137	.167	.000	.156	.000	.130	.003	.188	.000	.133	.003
	PAC_8	-.073	.345	-.057	.269	-.095	.009	-.046	.368	-.094	.012	-.060	.253
	PACDEE	-.178	.519	-.337	.060	.190	.193	.189	.344	.297	.051	.214	.298
	PACFLEI	.074	.829	.015	.932	-.028	.824	-.166	.382	-.200	.145	-.150	.421
	PACDFR	.073	.070	.092	.000	.046	.010	.053	.037	.085	.000	.052	.036
	PACFD	-.018	.789	-.080	.071	-.104	.001	-.102	.020	-.054	.075	-.018	.671
	PACDPAI	-.294	.039	-.064	.476	.022	.715	.073	.395	-.127	.044	-.189	.031
	PACDLTI	.607	.000	.411	.000	.373	.000	.315	.000	.474	.000	.497	.000
	PACDTLE	-.155	.565	.093	.595	-.500	.000	-.469	.017	-.689	.000	-.637	.002
PACFLTI	.142	.683	.115	.533	.190	.148	.275	.159	.350	.013	.317	.098	

B – Logistic Constant

a. The reference category is Normal weight

b. This parameter is set to zero because it is redundant

FVCDVEG - Daily consumption - Other vegetables; FVCDTOT - Daily consumption - Total fruits and vegetables; FVCGTOT - Daily consumption - Total fruits and vegetables; PAC_7 - Walked to work or school - Past 3 months; PAC_8 - Bicycled to work or school - Past 3 months; PACDEE - Daily energy expenditure - Leisure physical activities; PACFLEI - Participant in leisure physical activity; PACDFR - Frequency - All leisure physical activity lasting more than 15 minutes; PACFD - Participant in daily leisure physical activity lasting more than 15 minutes; PACDPAI Leisure physical activity index; PACDLTI - Leisure and transportation physical activity index; PACDTLE - Daily energy expenditure - Transportation and leisure physical activity; PACFLTI - Participant in leisure or transportation physical activity

4.5 Mathematical Relationship Amongst Factors

Based on the statistical data analysis carried out, the main factors and variables for obesity and overweight were identified and their respective mathematical relationships obtained. The mathematical formula indicates how the relevant factors contributing to the emergence of overweight and obesity. Table 4.5 shows the main *factors* and *variables* found to contribute to obesity. For example, the factor, fruit and vegetable consumption (FVC), has three variables (*daily consumption of other vegetables* (FVCDVEG), *daily consumption of total fruits and vegetables* (FVCDTOT), and *daily consumption of total fruits and vegetables* (FVCGTOT)). Then, the variable *daily consumption of other vegetables* (FVCDVEG) has two different answers (behaviors) which are less than 5 times (Answer 1) or 5 times and more (Answer 2).

Table 4.5 Main factors for the obesity model

Factors	Variable	Codes	Answers/Behaviors		
FVC: Fruit and vegetable consumption	Daily consumption - Other vegetables	FVCDVEG	1 Less than 5 times	2 5 times or more	
	Daily consumption - Total fruits and vegetables	FVCDTOT	1 Less than 5 times	2 5 times or more	
	Daily consumption - Total fruits and vegetables	FVCGTOT	1 Eats fruits and vegetables less than 5 times per day	2 Eats fruits and vegetables between 5 and 10 times per day	3 Eats fruits and vegetables more than 10 times per day
PAC: Physical activities	Walked to work or school - Past 3 months	PAC_7	1 Yes	2 No	
	Daily energy expenditure - Leisure physical activities	PACDEE	1 Less than 5 times	2 5 times or more	
	Participant in leisure physical activity	PACFLEI	1 Yes	2 No	
	Frequency - All leisure physical activity lasting more than 15 minutes	PACDFR	1 Regular	2 Occasional	3 Infrequent
	Participant in daily leisure physical activity lasting more than 15 minutes	PACFD	1 Yes	2 No	
	Leisure physical activity index	PACDPAI	1 Active	2 Moderate active	3 Inactive
	Leisure and transportation physical activity index	PACDLTI	1 Active	2 Moderate active	3 Inactive
	Daily energy expenditure - Transportation and leisure physical activity	PACDTLE	1 Less than 5 times	2 5 times or more	
	Participant in leisure or transportation physical activity	PACFLTI	1 Yes	2 No	

One advantage of logistic regression analysis is its ability to directly estimate the probability of the occurrence of an event. In the case of the dependent variable Y with only two possible states (1 or 0) and a set of the independent variables, x_1, x_2, \dots, x_P , the multiple logistic regression function can be written as [POH04]:

$$\text{logit}(p) (Y = 1) = \frac{1}{1+e^{-g(x)}}, g(x) = B_0 + B_1x_1 + \dots + B_P x_P \quad (\text{Equation 2})$$

To determine the equation for each weight category, the constants “ B_n ” of the regression function “ $g(x) = B_0 + B_1x_1 + \dots + B_P x_P$ ”, should be replaced by the factor’s variables’ logistic constants obtained from the *Multinomial Logistic Regression* analysis as summarized in Table 4.4; where B_0 equals the “intercept” value.

For example, for the category “Obese in 2009-2010” the equation is

$$g(x) = -0.836 - 0.211x_1 - 0.477x_2 + 0.193x_3 + 0.098x_4 - 0.073x_5 - 0.178x_6 + 0.074x_7 + 0.073x_8 - 0.018x_9 - 0.294x_{10} + 0.607x_{11} - 0.155x_{12} + 0.142x_{13} \quad (\text{Equation 3})$$

The standard interpretation of the logistic constant is that for a unit change in the independent variable (x_1, x_2, \dots, x_{13}), the logit outcome relative to the reference group (normal weight) is expected to change by its respective parameter estimate if all the other variables are held fixed [UCL19]. For instance, if the agent’s daily consumption of other vegetables (x_1) increases by one unit, the risk of the occurrence of obesity to normal weight would be expected to decrease by 0.069 while holding all other variables in the model constant. Intercept (B_0) is the multinomial logit estimate for obesity relative to normal weight when the independent variables in the model are evaluated at zero. If $x_1, x_2, \dots, x_{13} = 0$, and the intercept is simply the expected mean value of Y at that value (in this case, -0.836).

As the absolute value of the constant “ B_n ” is larger, its influence is larger too (positively or negatively). For example, equation 3 implies that the obese state is affected more strongly by “daily frequency in transportation and leisure physical activity (x_{12})” which has the biggest absolute value among the constants. In contrast, the obese state is affected less strongly by “participant in daily leisure physical activity lasting more than 15 minutes (x_9)” which has the smallest absolute value among the constants.

The next step is to build the CAS-based obesity model by using the set of equations obtained for the four weight statuses (underweight, normal weight, overweight and obese). First, I need to set the BMI status of agents representing a population. In 2009-2010, the ratio of underweight, normal weight, overweight and obese is 1.79%, 43.23%, 35.22%, and 19.76% respectively. I will use this ratio as an initial value of the BMI status in the population. Secondly, I will set rules and factors for the agents. Each agent has 13 different obesity factors. To define the rules in the time series for the agent, I will use the obesogenic state equations each year in the period. For example, in 2009-2010, I can set the agents in the obese state by using Equation 3. Thus, agents have four different equations (underweight, normal weight, overweight and obese) available each year for calculating their obesogenic state. The original dataset contains the person's perception of their obesogenic state. This information is used to guide a suitable equation for an individual.

4.6 Year-to-year analysis

Each of the agents (individuals) has different conditions on factor values (Tables 4.6, 4.7). Thus, with the computer model using the equations and datasets, I will carry out a year-to-year statistical data analysis simulation of people in a community. In other words, I will evaluate

people changing behaviors in terms of the relevant obesity factors identified and see how obesogenic state rates arise in the community as a consequence of people's actions (Figure 3.1).

Every year, an individual will have a transition probability for all their risk factors to move within the states. For example, in the year 2009-2010 and 2010, Daily consumption (FVCDVEG) has two statuses, 1 (less than 5 times) and 2 (5 times or more) for each obese status. The ratio changes in each factor between consecutive years are calculated and this will be applied to the model.

Table 4.6 FVCDVEG year to year analysis for 2009-2010

2009-2010				
HWT/FVCDVEG	less than 5 times 1	5 times or more 2	ratio 1	ratio 2
Underweight 1	290	7	0.976431	0.023569
Normal weight 2	6968	199	0.972234	0.027766
Overweight 3	5731	107	0.981672	0.018328
Obese 4	3220	55	0.983206	0.016794

Table 4.7 FVCDVEG year to year analysis for 2010

2010				
HWT/FVCDVEG	less than 5 times 1	5 times or more 2	ratio 1	ratio 2
Underweight 1	795	20	0.974326	0.025674
Normal weight 2	16698	499	0.970983	0.029017
Overweight 3	13849	280	0.980183	0.019817
Obese 4	7788	147	0.981474	0.018526

4.7 Obesity risk factor selection

Risk factors are selected based on the result of previous data analysis. Table 4.8 represents the risk factors that are used for the obesity model. The main risk factors are as follows:

- Fruit and vegetable consumption
 - FVCDVEG - Daily consumption - Other vegetables
 - FVCDTOT - Daily consumption - Total fruits and vegetables
- Physical activities
 - PAC_7 - Walked to work or school - Past 3 months
 - PACDEE - Daily energy expenditure - Leisure physical activities
 - PACFLEI - Participant in leisure physical activity
 - PACFD - Participant in daily leisure physical activity lasting more than 15 minutes
 - PACFLTI - Participant in leisure or transportation physical activity

Table 4.8 Risk factors for Obesity model

Risk Factor Code	Meaning	NetLogo element	Value range
FVCDVEG	Daily consumption of other vegetables	Switch(ON/OFF)	1 = Less than 5 times 2 = 5 times or more
FVCDTOT	Daily consumption of total fruits and vegetables	Switch(ON/OFF)	1 = Less than 5 times 2 = 5 times or more
PAC_7	Walked to work or school past 3 months	Switch(ON/OFF)	1 = Yes 2 = No
PACDEE	Daily energy expenditure leisure physical activities	Switch(ON/OFF)	1 = Less than 5 times 2 = 5 times or more
PACFLEI	Participant in leisure physical activity	Switch(ON/OFF)	1 = Yes 2 = No
PACFD	Participant in daily leisure physical activity lasting more than 15 minutes	Switch(ON/OFF)	1 = Yes 2 = No
PACFLTI	Participant in leisure or transportation physical activity	Switch(ON/OFF)	1 = Yes 2 = No

Chapter 5.

Model Implementation

Chapter 3 presents the conceptual design for this model, and Chapter 4 describes what kinds of data are used and how data analysis is implemented. Based on the previous chapters, I present the implemented obesity model by using the NetLogo multi-agent model.

5.1 Obesity model development process

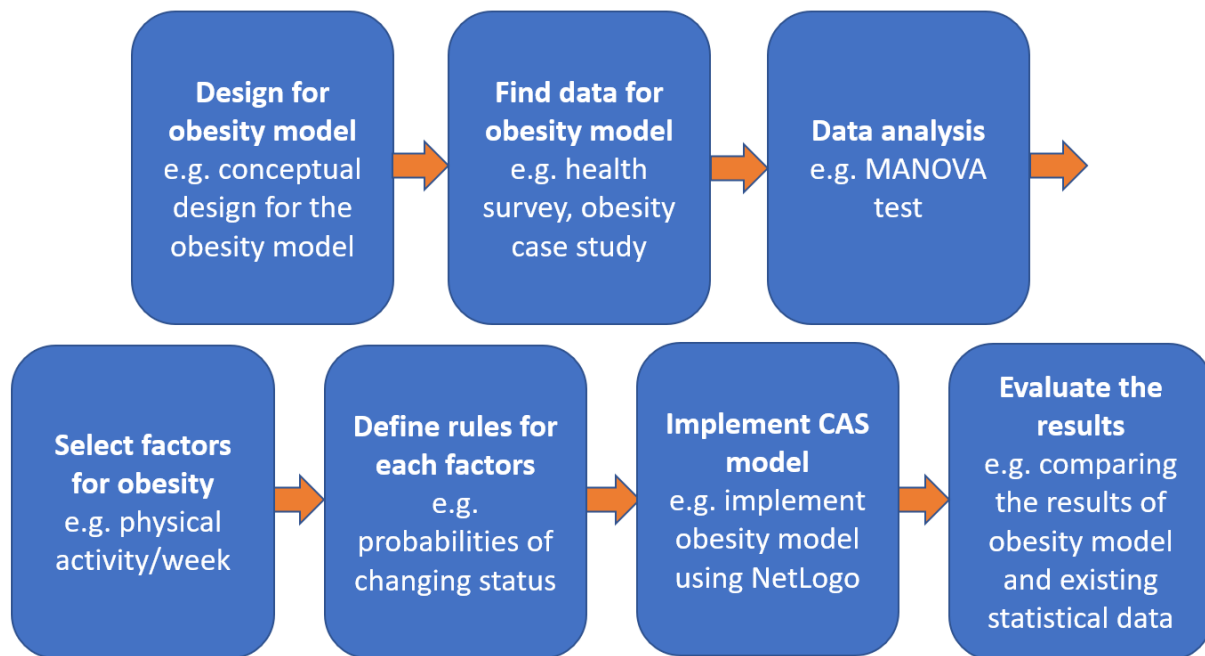


Figure 5.1 Implementation steps for the obesity model

The implementation process for the obesity model is following these 7 steps:

- 1) Design for obesity model: A conceptual design is conducted to build the obesity model.

2) Find data for the obesity model from various studies, such as health surveys, obesity case studies, etc. For this study, I am using data from the Canadian Community Health Survey (CCHS) on the NESSTAR data portal from Statistics Canada.

3) Data analysis: I implemented statistical data analysis such as the Normal Test, MANOVA Test, Multinomial Logistic Regression Test on the CCHS dataset to find out the interrelationship among the factors. These tests were performed using SPSS. Also, I implemented a year-to-year analysis that can show how people change their obesity status each year.

4) Select factors for obesity: Based on the data analysis results, I choose the factors for the obesity model.

5) Define rules for each factor: Based on steps 2 to 4, I defined rules (in this case, the probability of people's change status on each factor). For example, in the year 2009-2010 and 2010, the probability of changing the status of the obese group's daily consumption of other vegetables (FVCDVEG) 1 to 2 is 0.000198 for each factor.

6) Implement the CAS model: I implemented the obesity model using NetLogo.

7) Evaluate the results: I compared the result of the obesity model and existing statistical data.

Step 1, the design for the obesity model is outlined in detail in Chapter 3. step 2 to step 4 are described in Chapter 4. step 5 is explained in this Chapter, and step 6 will be elaborated on in Chapter 6.

5.2 Architecture of the obesity model implementation

The architecture of obesity model implementation consists of three main parts: population module, risk factor module, and the result module.

In the obesity modeling, I mainly used the IF ELSE statement. For example, if Agent 1's daily consumption of other vegetables (FVCDVEG) is smaller than the probability of FVCDVEG, then Agent 1's FVCDVEG status is 1 (this means agent 1 consumes vegetables less than 5 times a week) if not (ELSE), 2 (this means agent 1 consumes vegetables 5 times or more a week). The probability of FVCDVEG status is based on a year-to-year analysis in Chapter 4.

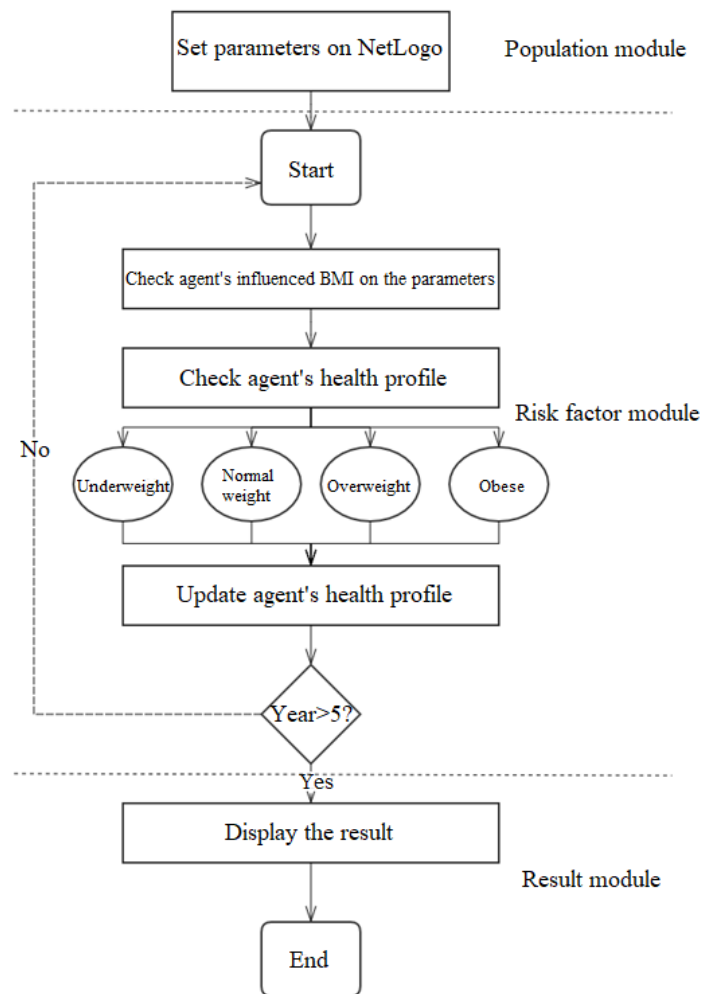


Figure 5.2 Flow chart for the obesity model

Figure 5.2 flow chart for the obesity model shows the overall structure of the obesity model implemented on NetLogo. A more detailed explanation follows.

5.2.1. Population module

The population module is implemented as ‘*to initial-setup*’ procedures. This procedure is to generate the baseline characteristic of the hypothetical person to match the data from health Canada. Each agent (person) gets aged year by year. Then the result is compared with the test set data.

5.2.2. Risk factor module

The risk factor module is used for estimating the risk of obesity based on the interaction among the risk factors. Each risk factor is in one of two statuses. For example, daily consumption of other vegetables (FVCDVEG) has two statuses such as 1 (Less than 5 times) and 2 (5 times or more). The probability of changing status is described in Chapter 3.

to check-health-status – This procedure is used to check the agent’s influenced BMI based on the parameter’s settings on NetLogo.

to health-status-update – This procedure is used to update each agent’s obesity status each year.

5.2.3. Result module

The results module is used to display and track each person(agent)'s obesity status change in the simulation.

to display_obesity_status – This procedure is used to display the overall obesity status result on the model.

to do-plots – This procedure is used to display the result of the agent’s health status update graph.

The following parameters and controls are used to carry out various functions such as selecting the main risk factors for simulating the obesity model.

Switch

This switch is used for the user to select the risk factors. For example, if the user wants to select the risk factor daily consumption of other vegetables (FVCDVEG), select FVCDVEG as on.

Slider

The slider is used for setting a percentage that indicates how much the individual has the risk factor value.

Button

The setup button is used for initializing and clearing all previous statuses. The go button is used to run the procedures and get a result of the agent’s health status.

Plot

This shows the obesity status in the form of a graph.

Chapter 6.

Performance Evaluation

This chapter provides the obesity model experimental results. First, a model calibration was executed to validate this obesity model. Second, several experiments were conducted to find out what are the major risk factors for obesity in various scenarios.

6.1 Evaluation method

The obesity model is generated based on the data analysis from the Canadian Community Health Survey (CCHS) on the NESSTAR data portal (Statistics Canada) which is described in Chapter 3. The empirical data from the CCHS is collected for 5 years (2009-2014). Therefore, the data of CCHS from 2009 to 2014 and the result of the obesity model which is generated in the 5 years of simulation are compared.

The total number of cases from 2009-2010 is 10,934. The data is split into two groups, train set, and test set for an 80/20 ratio. Therefore, the data is split into 8,751, and 2,183 cases respectively. Then, the dataset is classified for the independent variables (i.e., risk factors such as FVCDVEG, PAC_7, etc.) as X and the target variable as Y (i.e., Obesity status). Figure 6.1 represents the data set split process. First, the dataset is randomly divided into a train set and a test set. Then the dataset is separated into rows X and Y.

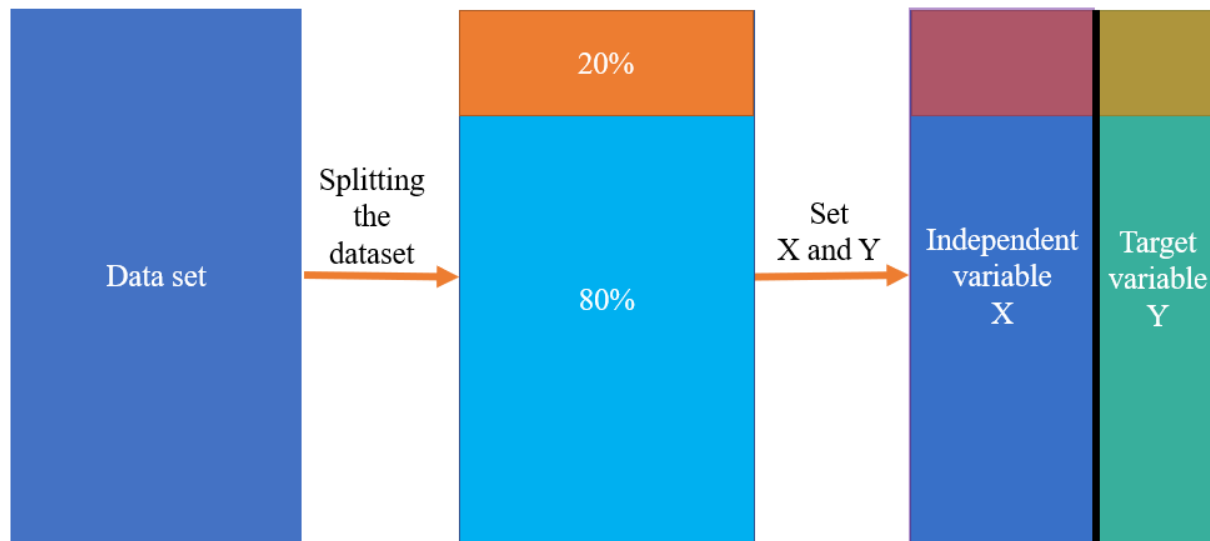


Figure 6.1 Dataset split process

6.2 Model calibration and validation

The model calibration process is adjusting and modifying the parameters to fit the model output and real results (i.e., empirical data from the CCHS 2009-2014). This can provide a good description of the system behavior of obesity and can also be obtained by confronting simulation results with actual data [JUD10] [WIL15]. Initially, the obesity model is built based on the data analysis in Chapter 4. Each risk factor and simple rules (probability of transferring their behavior and obesity state) are selected. Then, the dataset is divided into training data and testing data.

Table 6.1 Test result

Obese status	Actual data	Model test result	Standard deviation	Min	Max	Percentage error (%)
Underweight	18.44081	17	0	17	17	4.999068
Normal-weight	418.7103	424.8	43.12424	348	449	0.242302
Overweight	342.9066	344.2	4.868265	338	349	0.292725
Obese	219.9423	196	0	196	196	0.405171

Model calibration is executed to fit the actual data to the model test result. The model is started with the initial BMI equation given in Chapter 4. Data Analysis and then modified to influence the BMI equation to fit the actual data set. The table 6.1 shows the test results. First of all, the total population is set to 1,000. The numbers in actual data mean the number of people for each obese group, and the data is based on CCHS 2014. For example, the number 18.44081 in underweight represents that the number of people in the underweight group is 18.44081 out of 1,000 in the whole group. Secondly, the numbers in the model test result mean the average value of the results of running the obesity model simulation 30 times. For example, the number 344.2 in overweight indicates that the number of people in the overweight group is 344.2 out of 1,000, and its standard deviation is 4.868265. Lastly, as a percentage of true value and the values of the percentage error in the model test result are less than 5%, the percentage error is the difference between the measured value (simulated result) and the true value (actual result).

6.3 Model Experiment Sets

6.3.1 Experiment scenario 1: What are the risk factors for increasing obesity? Comparison with each risk independently

In this experiment set, I choose the specific risk factor to see how the risk factor impacts the model. One risk factor is selected and simulated 30 times, averaged, and compared to an unadjusted result. The unadjusted result means the baseline model test result which is calibrated with including the entire risk factors.

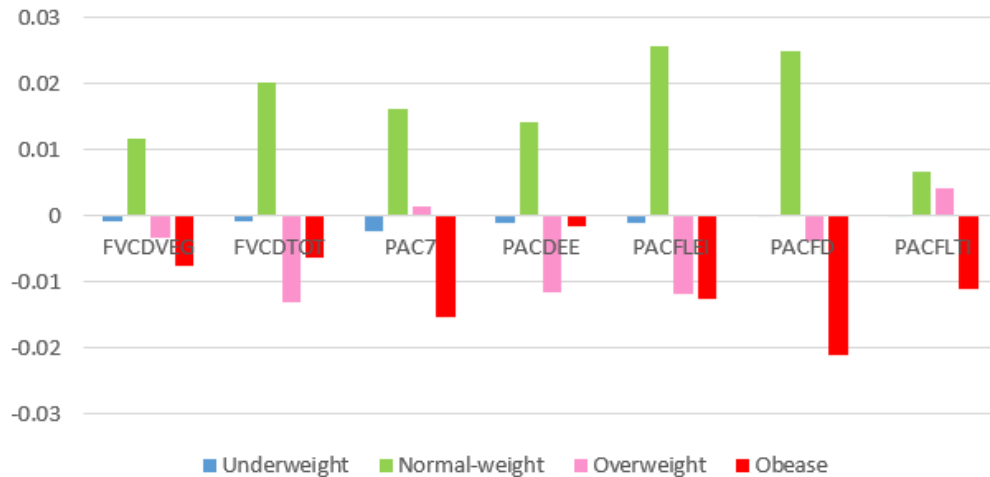


Figure 6.2 The influence of the risk factor in the obesity model

Figure 6.2 shows how each risk factor impacts each obese group. In this experiment, the result shows that the normal-weight group has distinctive characteristics compared to other groups because all risk factors in a normal-weight group have positive values. On the other hand, the rest of the groups generally shows negative values. For example, the normal-weight group in FVCDVEG has a positive value of 0.0116, whereas the underweight, overweight, and obese group have a negative value of 0.0008, 0.0033, 0.0075 respectively.

6.3.2 Experiment scenario 2: What are the risk factors for increasing obesity? Comparison between food consumption and physical activities factors

In this experiment, I tested what is the major risk factor to reduce obesity rates as a group. In this obesity model, to see what kinds of factors the major risk factors among other risk factors in obesity are, users can select and manipulate the intensity of risk factors such as food consumptions (FVCDVEG, FVCDTOT) and physical activities (PAC7, PACDEE, PACFLEI,

PACFD, PACFLTI). Therefore, multiple risk factors (food consumption and physical activity) were selected as major risk factors to reduce the risk of obesity. First, physical activity risk factors (PAC7, PACDEE, PACFLEI, PACFD, PACFLTI) are selected, and food consumption factors are excluded. In this test, the simulation is conducted 30 times, averaged, and compared to the unadjusted experiment. Contrary to expectations, the ratio of the overweight group increased by 0.6%, and the ratio of the obese group increased by 0.3% compared with previous test results. Therefore, in the next experiment, I selected fewer risk factors such as PAC7, PACFLEI, PACFLTI to see in detail. Compared with the unadjusted group, the ratio of the overweight group increased by 0.2%. However, the ratio of the obese group decreased by 1.6%.

In the next experiment, the risk factors which are in food consumption (FVCDVEG, FVCDTOT) were selected. The ratio of the obese group increased by 1.9%, and the ratio of other groups decreased.

6.3.3 Experiment scenario 3: What are the risk factors for increasing obesity? Combination of risk factors (food consumption and physical activities)

Based on experimental sets scenario 1 and scenario 2 results, now multiple risk factors were selected to get some insights when I combine risk factors;

In the first case, FVCDVEG, FVCDTOT (food consumption), PAC7, PACFLEI, and PACFLTI (physical activities) were selected. This case increased by 0.8% of the overweight rate but the obese rate reduced by 0.5%.

In the second case, FVCDVEG (food consumption), PAC7, PACFLEI, and PACFLTI (physical activities) were selected. The normal rate decreased by 2.6% but the overweight rate increased by 1.5%, and the obese rate increased by 1%.

In the last case, FVCDTOT (food consumption), PAC7, PACFLEI, and PACFLTI (physical activities) were selected. The normal rate decreased 1.6%, and the overweight rate increased by 1.1%.

Among the three sets of scenarios, a single risk factor, PACFLTI (Participant in leisure or transportation physical activity) impacts the most in the obesity status. In particular, it decreases the ratio of the overweight group by 0.4%. Also, combined risk factors, FVCDVEG, FVCDTOT, PAC7, PACFLEI, and PACFLTI can reduce the ratio of the obese group by 0.5%.

The most important risk factors in each obesity group are identified based on the three sets of experimental scenario results. With this information, at the individual level, each individual can find what kinds of strategies are the best fit to improve their current health status. Also, at a government or community level, more indoor and outdoor activities can be promoted to the public so that people have more chances to participate in leisure or physical activities. This can significantly improve physical health. Also, the government should increase access to nutritious foods and a healthier food choice policy for the public. Consuming unhealthy foods such as processed foods make people get high calories, high salt, and high sugar consumption. These are major leading risks of obesity [BAC19]. If governments and policymakers can promote these to the public properly, this will reduce the obesity rates and eventually reduce healthcare expenditure related to obesity such as diabetes, asthma, and heart disease.

Chapter 7.

Conclusion and Future Work

7.1 Summary of the Study

This study aims to build an obesity ABM model and find out how obesity is affected by risk factors. I approach the topic through three objectives. The first object is to find out the risk factors that affect obesity. The second objective is to understand the complex interdependencies among risk factors. The last objective is to develop an agent-based model that can simulate the obesity rate based on statistical analysis and to see the interaction of the people (agents) affects on the society (system).

I followed 7 steps to implement the obesity model: 1) Designing for obesity model, 2) Finding data for the obesity model from various studies, such as health surveys and obesity case studies,3) Data analysis, 4) Selecting factors for obesity, 5) Defining rules for each factor, 6) Implementing the CAS model, and 7) Evaluating the results.

The obesity model consists of three key elements: agent, risk factors, and time. In detail, in obesity modeling an agent can represent each individual in the population and each person's characteristics are set as risk factors (attributes of agents) and set rules for changing their status over time to see the dynamics of obesity.

The data used for this study was provided by Statistics Canada, and the Canadian Community Health Survey (CCHS), available on the NESSTAR data portal (Statistics Canada).

This survey is a cross-sectional survey that collects self-reported information related to health status, health care utilization, and health determinants for the Canadian population. It relies upon a large sample of respondents and is designed to provide reliable health estimates at the regional level [STA18].

To build an ABM, it is necessary to find out which risk factors are associated with obesity and to what extent they act. As obesity is a complex disorder, it is assumed that various risk factors influence obesity, and the risk factors interact with one another.

I obtain data from various health-related risk factors such as chronic conditions, income, age, and sex. Also, each risk factor has sub-factors, but I only want to include risk factors truly associated with obesity that have been identified by statistical data analysis.

As the main idea is to try to create a mathematical formula to use in the construction of the modeling, it is verified that the best statistical analysis, for the database of those years and subsequent use for the creation of the software, would be to perform the Normal Test, MANOVA and Multivariate Logistic Regression, as described in more detail in Chapter 4 .

Based on the statistical data analysis carried out, the main factors and variables for obesity and overweight were identified and their respective mathematical relationships obtained. The mathematical formula indicates how the relevant factors contribute to the emergence of overweight and obesity.

The architecture of the obesity model implementation consists of three main parts: the population module, the risk factor module, and the results module. In this modeling, I mainly used the IF ELSE statement. First, the population module is used for loading data from text files

and sets parameters on the model. Second, the risk factor module is used for estimating the risk of obesity based on the interaction among the risk factors. Each agent has risk factors and obesity status. A risk factor is in one of two states, and obesity has four statuses. The probability of changing status is described in Chapter 4. Based on the transition probabilities, I calculate the agent's health status. Then, the agent's health profile is saved. Lastly, the results module is used to display and track each person (agent)'s obesity status change in the simulation.

The parameters and controls to carry out various functions for simulating the obesity model are as follows: switch, slider, set-up button, load file Button, and the go button. The switch is used for the user to select the risk factors. The slider is used for setting a percentage that shows how much value the individual has for the risk factor. The setup button is used for initializing and clearing all previous statuses. The load file button is used to read input experimental data in the obesity model from a text file. The go button is used for running and getting results of the agents' health status.

Performance evaluation was conducted to examine whether the obesity model can simulate the obesity rate based on statistical analysis. For this evaluation, the data of CCHS from 2009 to 2014 and the result of the obesity model which is generated in the 5 years of simulation are compared. Model calibration was executed to fit the actual data to the model test result. My model starts with the initial BMI equation based on Chapter 4. Data Analysis was then modified to influence the BMI equation to fit the actual data set.

The result of the model test shows that the percentage error is less than 5% as presented in Table 6.1. This means that the obesity model has high validity in predicting obesity for each risk factor. With this information, at the individual level, each individual can find what kinds of

strategies are best fit to improve their current health status. Also, at a government or community level, more indoor and outdoor activities can be promoted to the public so that people have more chances to participate in leisure or physical activities. This can significantly improve physical health. Also, the government should increase access to nutritious foods and a healthier food choice policy for the public.

7.2. Strengths of the CAS-based Obesity Model

1. Agent-based modeling (ABM) is a powerful simulation modeling method that can be applied to real problems. In this study, a variety of risk factors affecting obesity and their inter-relationships are considered by employing statistical analysis from a complex adaptive system perspective. This model can potentially assist to save some associated costs from obesity by helping to prevent and improve obesity management. For example, it can be useful in improving obesity conditions in adolescents. According to a recent WHO report, the rate of obesity increase in adolescents is severe, but unlike adults, the burden on obesity in the country or society is very large because it continues to burden the national cost for a long time. Thus, this model can potentially assist to save some associated costs from obesity by improving adolescent obesity management.

2. Unlike the previous studies which consider only one decisive risk factor, my study is comprehensive as it includes a large range of different risk factors. My study is also applying correlations amongst the risk factors into a CAS-based obesity model. Thus, in this model, users can manipulate various risk factors that influence obesity. This makes it possible to observe which risk factors change the obesity status. Consequently, it is possible to prepare an appropriate response to various obesity situations using this obesity model.

3. In ABM, a system is modeled as a collection of autonomous decision-making entities called agents. In this model, each agent represents human individuals and interacts with the other agents. Each agent evaluates the situation individually and makes a decision based on a set of rules. Therefore, this obesity model can be useful for finding out what risk factors are affecting each agent's(individual's) obesity status. From a micro level, an individual can obtain a method to reduce or prevent obesity that is appropriate for their health condition. Moreover, from a macro level, this model enables governments to legislate effective policies to reduce and prevent obesity.

7.3 The Limitations of Obesity Model and Future work

Regardless of these advantages, there are some limitations of the developed model of this study. Therefore, follow-up studies to compensate for these limitations will continue to take place.

1. As the subjects of this study are limited to adults in Canada, the bias of the data may be raised. The reason is that risk factors can vary depending on region, age, race, and culture, and these differences can affect lifestyle and social relationships and can result in different interactions of risk factors. For example, some regions may not be able to eat less fruit or vegetables in a certain season or may not be able to eat certain foods depending on religion or social practices. In addition, women may have limited social activities based on social practices.

2. The number of subjects of this study is large, but some limitations are bound to predict only the variable relationship at a limited time because it is fixed to a specific time (from 2009 to 2014). Therefore, there is a complementary need for longitudinal studies that can repeatedly

investigate the characteristics of the same subject over time and analyze the changes or factors that affect them.

3. In this study, obesity-related risk factors were selected based on the findings of the existing research datasets. Therefore, there may be risk factors that have not been selected, which requires continued research and evaluation.

References

- [ABA17] Abarca-Gómez, L., Abdeen, Z. A., Hamid, Z. A., Abu-Rmeileh, N. M., Acosta-Cazares, B., Acuin, C., ... & Agyemang, C. (2017). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128· 9 million children, adolescents, and adults. *The Lancet*, 390(10113), 2627-2642.
- [AZI16] Aziza, R., Borgi, A., Zgaya, H., & Guinhouya, B. (2016). SimNCD: An agent-based formalism for the study of noncommunicable diseases. *Engineering Applications of Artificial Intelligence*, 52, 235-247.
- [BAC19] Bacon, S. L., Campbell, N. R., Raine, K. D., Tsuyuki, R. T., Khan, N. A., Arango, M., & Kaczorowski, J. (2019). Canada's new Healthy Eating Strategy: Implications for health care professionals and a call to action. *Canadian Pharmacists Journal/Revue des Pharmaciens du Canada*, 152(3), 151-157.
- [BEN13] Bendsen, N. T., Christensen, R., Bartels, E. M., Kok, F. J., Sierksma, A., Raben, A., & Astrup, A. (2013). Is beer consumption related to measures of abdominal and general obesity? A systematic review and meta-analysis. *Nutrition reviews*, 71(2), 67-87.
- [BEW03] Bewick, V., Cheek, L., & Ball, J. (2003). Statistics review 7: Correlation and regression. *Critical care*, 7(6), 1-9.
- [BON02] Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the national academy of sciences*, 99(suppl 3), 7280-7287.

- [BRU15] Bruch, E., & Atwell, J. (2015). Agent-based models in empirical social research. *Sociological methods & research*, 44(2), 186-221.
- [CEN18] Centers for Disease Control and Prevention. Adult Obesity Causes & Consequences. Available from: <https://www.cdc.gov/obesity/adult/causes.html>
- [CHA14] Charlton, K., Kowal, P., Soriano, M. M., Williams, S., Banks, E., Vo, K., & Byles, J. (2014). Fruit and vegetable intake and body mass index in a large sample of middle-aged Australian men and women. *Nutrients*, 6(6), 2305-2319.
- [CHR16] Christiansen, S. C., Schatz, M., Yang, S. J., Ngor, E., Chen, W., & Zuraw, B. L. (2016). Hypertension and asthma: a comorbid relationship. *The Journal of Allergy and Clinical Immunology: In Practice*, 4(1), 76-81.
- [COD15] Codella, J., Safdar, N., Heffernan, R., & Alagoz, O. (2015). An agent-based simulation model for Clostridium difficile infection control. *Medical Decision Making*, 35(2), 211-229.
- [COL18] Colapinto, C. K., Graham, J., & St-Pierre, S. (2018). Trends and correlates of frequency of fruit and vegetable consumption, 2007 to 2014. *Health Rep*, 29(1), 9-14.
- [DAR15] Dare, S., Mackay, D. F., & Pell, J. P. (2015). Relationship between smoking and obesity: a cross-sectional study of 499,504 middle-aged adults in the UK general population. *PloS one*, 10(4), e0123579.
- [EPS95] Epstein, L. H., Valoski, A. M., Vara, L. S., McCurley, J., Wisniewski, L., Kalarchian, M. A., ... & Shrager, L. R. (1995). Effects of decreasing sedentary behavior and increasing activity on weight change in obese children. *Health psychology*, 14(2), 109.

[EUB04] Eubank, S., Guclu, H., Kumar, V. A., Marathe, M. V., Srinivasan, A., Toroczkai, Z., & Wang, N. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988), 180-184.

[FRE08] French, A., Macedo, M., Poulsen, J., Waterson, T., & Yu, A. (2008). Multivariate analysis of variance (MANOVA).

[FUC02] Fuchs, C., & Hofkirchner, W. (2002). Information in social systems. *INTERNATIONAL CONGRESS OF THE IASS-AIS* (Vol. 7).

[HEN16] Hennessy, E., Ornstein, J. T., Economos, C. D., Herzog, J. B., Lynskey, V., Coffield, E., & Hammond, R. A. (2016). Peer reviewed: designing an agent-based model for childhood obesity interventions: a case study of ChildObesity180. *Preventing chronic disease*, 13.

[HIL11] Hills, A. P., Andersen, L. B., & Byrne, N. M. (2011). Physical activity and obesity in children. *British journal of sports medicine*, 45(11), 866-870.

[HOU15] Hough, G., & Sosa, M. (2015). Food choice in low income populations—A review. *Food quality and preference*, 40, 334-342.

[IBM19] IBM SPSS Statistics. IBM SPSS software. Available from:

<https://www.ibm.com/analytics/spss-statistics-software>

[JAN13] Janssen, I. (2013). The public health burden of obesity in Canada. *Canadian Journal of Diabetes*, 37(2), 90-96.

[JUD10] Judd, S. (2010). *The MBR book: principles and applications of membrane bioreactors for water and wastewater treatment*. Elsevier.

- [LEE13] Lee, B. Y., Wong, K. F., Bartsch, S. M., Yilmaz, S. L., Avery, T. R., Brown, S. T., ... & Huang, S. S. (2013). The Regional Healthcare Ecosystem Analyst (RHEA): a simulation modeling tool to assist infectious disease control in a health system. *Journal of the American Medical Informatics Association*, 20(e1), e139-e146.
- [LIC00] Lichtenstein, B. B. (2000). The matrix of complexity: A multi-disciplinary approach for studying emergence in coevolution. *Online retrieved on January, 20, 2014.*
- [MAL16] Malta, D. C., Santos, M. A. S., Andrade, S. S. C. D. A., Oliveira, T. P., Stopa, S. R., Oliveira, M. M. D., & Jaime, P. (2016). Tendência temporal dos indicadores de excesso de peso em adultos nas capitais brasileiras, 2006-2013. *Ciência & Saúde Coletiva*, 21, 1061-1069.
- [MAY19] Mayo Clinic. Obesity. Available from: <https://www.mayoclinic.org/diseases-conditions/obesity/symptoms-causes/syc-20375742>
- [MIL09] Miller, J. H., Page, S. E., & Page, S. (2009). *Complex adaptive systems*. Princeton university press.
- [MOK03] Mokdad, A. H., Ford, E. S., Bowman, B. A., Dietz, W. H., Vinicor, F., Bales, V. S., & Marks, J. S. (2003). Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. *Jama*, 289(1), 76-79.
- [MUS19] Must, A., Spadano, J., Coakley, E. H., Field, A. E., Colditz, G., & Dietz, W. H. (1999). The disease burden associated with overweight and obesity. *Jama*, 282(16), 1523-1529.
- [PAE02] Paeratakul, S., Lovejoy, J. C., Ryan, D. H., & Bray, G. A. (2002). The relation of gender, race and socioeconomic status to obesity and obesity comorbidities in a sample of US adults. *International journal of obesity*, 26(9), 1205-1210.

- [PAR11] Parker, J., & Epstein, J. M. (2011). A distributed platform for global-scale agent-based models of disease transmission. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 22(1), 1-25.
- [PER09] Perez, L., & Dragicevic, S. (2009). An agent-based approach for modeling dynamics of contagious disease spread. *International journal of health geographics*, 8(1), 1-17.
- [POH04] Pohar, M., Blas, M., & Turk, S. (2004). Comparison of logistic regression and linear discriminant analysis: a simulation study. *Metodoloski zvezki*, 1(1), 143.
- [PRA14] Pratt, L. A., & Brody, D. J. (2014). Depression and obesity in the U.S. adult household population, 2005-2010. *NCHS data brief*, (167), 1–8.
- [RED08] Reddy Thavanati, P. K., Kanala, K. R., De Dios, A. E., & Cantu Garza, J. M. (2008). Age-related correlation between antioxidant enzymes and DNA damage with smoking and body mass index. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 63(4), 360-364.
- [REM15] Remedy Health Media. Risk Factors and Causes of Obesity. Available from: <http://www.healthcommunities.com/obesity/causes.shtml>
- [ROB01] Robinson, T. N. (2001). Television viewing and childhood obesity. *Pediatric Clinics of North America*, 48(4), 1017-1025.
- [ROE10] Roelofs, R., Gurgel, R. Q., Wendte, J., Polderman, J., Barreto-Filho, J. A. S., Solé, D., ... & Agyemang, C. (2010). Relationship between asthma and high blood pressure among adolescents in Aracaju, Brazil. *Journal of Asthma*, 47(6), 639-643.

[SAY11] Sayon-Orea, C., Martinez-Gonzalez, M. A., & Bes-Rastrollo, M. (2011). Alcohol consumption and body weight: a systematic review. *Nutrition reviews*, 69(8), 419-431.

[STA18] Statistic Canada. (2018). *Canadian Community Health Survey 2014 - Annual Component (CCHS)*. Available from:

https://www23.statcan.gc.ca/imdb/p3Instr.pl?Function=assembleInstr&Item_Id=214314&TET=1

[STA19] Statistics Canada. (2019). *Canadian Community Health Survey (CCHS)*. NESSTAR data portal. Available from: <http://libproxy.uwinnipeg.ca/login?url=http://dli-idd-nesstar.statcan.gc.ca/webview/>

[TRA18] Tracy, M., Cerdá, M., & Keyes, K. M. (2018). Agent-based modeling in public health: current applications and future directions. *Annual review of public health*, 39, 77-94.

[TRE17] Tremblay, M. S., Aubert, S., Barnes, J. D., Saunders, T. J., Carson, V., Latimer-Cheung, A. E., ... & Chinapaw, M. J. (2017). Sedentary behavior research network (SBRN)—terminology consensus project process and outcome. *International journal of behavioral nutrition and physical activity*, 14(1), 1-17.

[UCL19] UCLA: Statistical Consulting Group. Multinomial logistic regression. Available from: <https://stats.idre.ucla.edu/spss/output/multinomial-logistic-regression/>

[USD16] US Department of Health and Human Services. What causes obesity & overweight? Available from: <https://www.nichd.nih.gov/health/topics/obesity/conditioninfo/cause>

[WHO14] World Health Organization. (2014). *Global status report on noncommunicable diseases 2014* (No. WHO/NMH/NVI/15.1). World Health Organization.

[WHO18] World Health Organization. (2018). Noncommunicable diseases country profiles 2014.

[WHO19] World Health Organization. (2019). Mean Body Mass Index (BMI). Available from: https://www.who.int/gho/ncd/risk_factors/bmi_text/en/

[WHO21] World Health Organization. (2021). Obesity and overweight fact-sheets. Available from: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>

[WIL15] Wilensky, U., & Rand, W. (2015). *An introduction to agent-based modeling: modeling natural, social, and engineered complex systems with NetLogo*. Mit Press.

[WOO09] Wooldridge, M. (2009). *An introduction to multiagent systems*. John Wiley & Sons.

[YEO10] Yeomans, M. R. (2010). Alcohol, appetite and energy balance: is alcohol intake a risk factor for obesity?. *Physiology & behavior*, 100(1), 82-89.