Prediction for Canadian Federal Election Aided by Canadian Community Health Survey

by

Qi Wen

B.Sc., East China Normal University, 2009

Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

SIMON FRASER UNIVERSITY
Fall 2019

Approval

Name:	**Qi Wen**
Degree:	**Master of Science (Statistics)**
Title:	**Prediction for Canadian Federal Election Aided by Canadian Community Health Survey**
Examining Committee:	**Chair:** Jinko Graham Professor
	X. Joan Hu Senior Supervisor Professor
	Derek Bingham Supervisor Professor
	Kennedy Stewart External Examiner Mayor of Vancouver Associate Professor School of Public Policy
Date Defended:	**September 05, 2019**

Abstract

This project aims to develop predictive models for Canadian federal elections. We begin with explanatory analyses of two sets of data: some publicly accessible election data and some extracted data from the Canadian Community Health Survey (CCHS) 2007-2018 on life satisfaction and other potentially associated social-demographics. We propose to predict for federal election outcomes using the information on longitudinal Canadian life satisfaction. Specifically, we model the federal election outcome for a riding in change from its previous election jointly with its longitudinal life satisfaction since the previous election. Election data from years 2008 and 2011 and the CCHS data of 2008-2011 are employed to fit the model via both the two-stage estimation and the maximum likelihood estimation by the Monte Carlo EM algorithm. The analysis results indicate that life satisfaction is an important factor in election prediction. It appears that young adults are more likely to vote for a change but male voters are less likely to do so. Using voter information or CCHS respondent's information to model the election outcomes produce different estimation results. Two applications of the proposed approach are presented to further illustrate the proposed joint modeling approach.

Keywords: General linear mixed-effects model; Joint modeling; Logistic regression model; Monte Carlo EM algorithm; Two-stage estimation.

Acknowledgements

First and foremost, I would like to thank my senior supervisor Dr. Joan Hu. Thank you Joan for taking a chance on the potential you saw in me, and for your endless encouragement and support over the past two years. I want to thank Dr. Derek Bingham and Dr. Kennedy Stewart for taking the time to review this project and serving on my committee. I want to extend a special thanks to Dr. Kennedy Stewart for giving me the opportunity to work for him and inspiring this project.

My gratitude also goes out to all of my professors here at Simon Fraser University for their exceptional lectures. Thank you to all my friends and fellow students for your support. My time here would not have been the same without all of you!

Lastly but not least, I would like to thank my family. I would not be where I am today without your unconditional love and constant words of encouragement.

Disclaimer: This research was supported by funds to the Canadian Research Data Centre Network (CRDCN) from the Social Sciences and Humanities Research Council (SSHRC), the Canadian Institute for Health Research (CIHR), the Canadian Foundation for Innovation (CFI), and Statistics Canada. Although the research and analysis are based on data from Statistics Canada, the opinions expressed do not represent the views of Statistics Canada.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background and Motivation

Canada is a representative democracy. In Canadian federal election, each eligible Canadian citizen has the right to vote for the candidate to represent a riding in the House of Commons. Federal elections are under the administration of Elections Canada, which is an independent and non-partisan agency. Based on geographical divisions, Canada is divided into constituencies, also called electoral districts or ridings. There are 338 ridings spread across the whole nation, each one of the ten provinces has multiple ridings, and each one of the three territories is one riding itself. Each riding corresponds to one seat in the House of Commons.

In every riding, any number of candidates can run for election either under a registered political party or as an independent candidate. The eligible citizens in each riding may vote for one of the candidates. Canada's electoral system is called a "single-member plurality" system. Under this system, the candidate who receives the highest number of votes wins the election in every riding. To be elected, the candidate does not need to receive an absolute majority of the votes but only need to receive votes more than any other candidate in the same riding. This elected candidate therefore wins a seat in the House of Commons and serves as this riding's Member of Parliament (MP).

If the majority of the MPs come from one political party, then the leader of that party will be appointed as the Prime Minister. This is called a majority government. If no political party has a majority of the seats, usually the party with the most seats will form the government, and this is called a minority government. For example, it requires at least 170 seats out of 338 seats to form a majority government. The creation of majority governments is one of the features of the single-member plurality system; a party that does not have the support of a majority of Canadians can still win the majority of the seats in the House of Commons and form a government (Ward, N. & Courtney, J., 2018).

There has been a surge of interest in forecasting election outcomes (Campbell, J. E., 1992). One dominant approach is to use the opinion polls conducted before the election to predict the upcoming election. Rosenthal (2011) proposed a simple seat projection model to predict the Canadian federal election results in year 2011. This model began in each riding with the results for each party from the previous election in the year 2008 and then was adjusted with the latest pre-election opinion polls. Recently, Tang (2017) compared the discrepancies between the real outcomes of recent Canadian federal elections and the seat predictions by the existing approaches that used the opinion polls to make seat projection, such as the estimation method proposed by Rosenthal (2011). The simulation study showed that the violation of the assumptions that were required by the estimation methods may impact the prediction results.

The other major approach is to use a regression model with social, economic, political, and demographic factors as the predictors to forecast election outcomes. Campbell (1992) proposed an Ordinary Least Square (OLS) regression model to forecast the election results in the United States. The predictors used in this model were the national economic status, the economic growth in the state, and some political factors such as the partisan division of the state legislature. Bélanger & Godbout (2010) also proposed an OLS regression model to predict Canadian federal election results. They chose to use three political and economic variables as the predictors: the unemployment rate, an indicator of overall satisfaction with the incumbent administration, and a measure of government longevity. They suggested that the voters generally believed the incumbent political parties were responsible for the country's economic status. A bad economy, indicated by high unemployment rate, would raise public dissatisfaction so that voters might choose to vote for one of the opposition parties. Both the unemployment rate and the indicator of satisfaction are overall measurements of the entire nation. We hypothesize that life satisfaction status among the voters could be used as one of the predictors in the regression model to forecast the federal election results.

The Canadian Community Health Survey (CCHS) is a national survey which provides us the information of life satisfaction among Canadians. Much research has been conducted using the CCHS data to explore life satisfaction among Canadians and its relation to other social-demographic factors. For example, Lombardo et al. (2018) concluded that subjective mental health was an important factor associated with life satisfaction. In addition, Liu et al. (2015) found that the life satisfaction level was affected by the subject's gender, age, marital status, education, health status, employment status, and household income. However, the life satisfaction status among Canadians provided by CCHS has never been linked to Canadian federal election outcomes.

1.2 Objectives

In the 2015 Canadian federal election, the Liberals had a big win while the Conservatives lost for the first-time since 2006. Clarke et al. (2017) discussed the factors that led to the winning of the Liberals, such as the bad economy, the high unemployment rate, the facts that the Conservatives were not popular and many voters did not believe that their leader, Steven Harper, could manage the country well.

Facing the bad economy, high unemployment rate, and an untrusted government, Canadians felt unhappy about their current lives (Clarke et al., 2017). Therefore they might tend to vote for a new government. We hypothesize that the "unhappiness" contributed to the big loss for Conservatives in the 2015 federal election. That is, we hypothesize that if people are happier or more satisfied with their lives, they will tend to vote for the incumbent, while if people are not happy or less satisfied with their lives, they will tend to vote for a change, i.e., another political party. We develop a joint modeling approach to predict Canadian federal election by including life satisfaction among Canadians. This project's objectives are as follows.

- Primary Objective: To develop predictive models for federal election outcomes, with particular attention to establishing them associated with Canadians' life satisfaction since the last election.

- Secondary Objectives:
  (i) To explore publicly accessible Canadian federal election data;
  (ii) To explore Canadian life satisfaction time trend using the CCHS data;
  (iii) To study the association of federal election outcome with Canadians' life satisfaction since the last election in the presence of a few selected socio-demographic variables.

1.3 Outline

We organize this project as follows. Chapter 2 starts by introducing Canadian federal election related data and showcasing some preliminary analysis. Chapter 3 introduces the second data source, the CCHS along with some preliminary analysis. We also make a comparison of these two data sources and show the practicability of using both data sources to predict the election outcomes. In Chapter 4, we introduce the notation, formulation, and the proposed joint modeling approach. Two different estimation methods, a two-stage estimation method, and a maximum likelihood estimation method are introduced, and their results are compared. Chapter 5 summarizes the prediction for the upcoming Canadian federal election and other application of the proposed joint models. Final remarks and an outlook for the future investigation concludes this project in Chapter 6.

We use SAS 9.4 (SAS Institute Inc., 2014) and R (R Core Team, 2017) to fit the proposed joint models and compute the analysis results. Particularly, for the two-stage estimation approach, we use the Proc Mixed procedure and the Proc Logistic procedure in SAS 9.4 to fit the joint models separately. For the maximum likelihood estimation approach, we use the built-in R function optim to maximize the likelihood function.

Chapter 2

Canadian Federal Elections

The first Canadian federal election was held in 1867. There have been 42 federal elections, with the most recent being held in 2015. The next federal election will be held in October, 2019. This chapter focuses on the most recent five elections (from the 38^{th} to the 42^{nd} election) held in the years 2004, 2006, 2008, 2011 and 2015. Section 2.1 begins by introducing some federal election administrative data, including the election results and voter turnout information. The descriptive and inferential analysis is also presented. Section 2.2 describes and summarizes the federal election survey data, including the opinion polls conducted before the election and the survey of electors conducted after the election.

2.1 Federal Election Administrative Data

Canada is divided into ridings based on geographical divisions and each riding has multiple polling stations. During the federal election, eligible electors go to the polling station to vote. An elector is a person who has the right to vote in the election. With very few exceptions, this person must be a Canadian citizen, at least 18 years old on election day and lives in the riding. To vote in a federal election, this person must be registered on the list of electors. The electors who have voted in the election are called voters. In this section, two types of administrative data of federal election are presented. The first one is the federal election results that summarize the votes information for each election. The second administrative data are the voter turnout data that summarize the information of the voters.

2.1.1 Federal Election Results

The Canadian federal election results are taken from the Elections Canada (2019). The results summarize the voting results of Canadian federal election and the data are publicly accessible. The following information is available for each riding in the data of federal election results.

- The population.

- The number of electors on the list.
- The number of polling stations.
- The number of valid votes and rejected votes.
- Every candidate's name and political affiliation.
- The incumbent's name and political affiliation.
- The elected MP's name and political affiliation.
- The number of votes each candidate received.

Preliminary Analysis

In the elections held in the year 2004, 2006, 2008 and 2011, there were 308 seats in the House of Commons, or there are 308 ridings across Canada. Due to the redistribution of electoral districts conducted in year 2012, the number of seats in the 2015 federal election was increased to 338 seats. At the time of the most recent federal election in 2015, there were 23 registered political parties in Canada, and the dominant parties have tended to be Liberals, Conservatives and New Democratic Party (NDP).

Year	**Election**	**Winning Party**	**Winning Seats**
2004	38th	Liberals	135/308
2006	39th	Conservatives	124/308
2008	40th	Conservatives	143/308
2011	41st	Conservatives	166/308
2015	42nd	Liberals	184/338
2019	43rd	?	?/338

Above is a summary of Canadian federal election results since year 2004. The Conservatives won the federal election three times in a row from year 2006 to year 2011, and the number of winning seat increased from 124 in year 2006 to 166 in year 2011, out of 308 seats. However in the most recent federal election in year 2015, the Conservatives lost the battle to the Liberals; the Liberals won 184 seats out of 338 seats and formed a majority government. The upcoming federal election will be held this year in October and the total number of seats is the same as the previous election in year 2015.

Table 1 and Figure 1 summarize the distribution of seats among the political parties for each election year. We summarize the four major political parties separately as four categories, the "Conservatives", the "Liberals", "NDP" and "Bloc Québécois". All the other political parties are grouped into the fifth category "Others". The blue line shows that the percentage of seats the Conservatives won keep increasing from election in year 2004 to year 2011 (32.14% vs. 53.90%) but has a huge drop in year 2015 (29.29%). The red line shows that the percentage of seats the Liberals won has a decreasing trend in the first four elections (43.83% in year 2004 vs. 11.04% in year 2011) but has a big jump in the last election in year 2015 (54.44%). The grey line shows that NDP has a similar changing pattern as the Conservative party has, the percentage of seat increases from the election in year 2004 to year 2011 (6.17% vs. 33.44%) but decreases in year 2015 (13.02%). As the yellow line shows, the percentage of seats that Bloc Québécois won decreases significantly over the decade (17.53% in year 2004 vs. 2.96% in year 2015). It is because this party is used to be the dominant political party in Quebec but since year 2011 many voters change to vote for the other three major political parties. The other political parties only have won one or two seats in the past five elections.

Figure 2 summarizes the percentage of votes that each political party received in the elections between year 2004 and 2015. Although the numbers of votes that each political party received do not determine the winning political party directly, these numbers could still show the popularity distribution among the parties. As this figure shows, the distribution of votes among these political parties has a similar pattern as the distribution of seats in Figure 1.

Figure 3 summarizes the distribution of whether or not the incumbent political party was re-elected for each election. That is, if the incumbent's political affiliation is re-elected in the current election then it is "Yes", otherwise it is "No". As the blue line shows, the number of incumbents re-elected decreased in the three elections (232 seats in year 2006, 217 seats in year 2008, and 207 seats in year 2011). Note that We are not able to summarize this information for election in year 2015 due to the redistribution of riding conducted before this election.

Figure 4 summarizes the distribution of elected MP's gender for each federal election. The percentage of elected female MPs shows an increasing trend over the decade (21.10% in year 2004 vs. 26.04% in year 2015). More and more females have been elected as the MPs, however the difference between male MPs and female MPs is still significant, for example in year 2015 which has the highest number of female MPs, there are only 88 female MPs but 205 male MPs.

2.1.2 Voter Turnout

Since 2004, under the Chief Electoral Officer's authority, Elections Canada has used the administrative data from the electoral process to obtain a sample of electors who voted for the election (Elections Canada, 2004). Along with the date of birth of each sampled elector obtained from the National Register of Electors, Election Canada produces the estimates of electors in population and the number of voters for each province and territory. For an election in year 2008 and after, these estimates are further divided into subgroups of age and gender. For the election in year 2004, only subgroups of age are provided.

The age as of polling day is calculated based on the date of birth and is divided into seven subgroups, 18-24 years old, 25-34 years old, 35-44 years old, 45-54 years old, 55-64 years old, 65-74 years old and 75 or above years old. Gender has two subgroups, male and female. Note that all the estimates are only provided for each province/territory but not available for each riding. For every province or territory, the following information is collected and summarized under each gender and age group combination.

- The number of registered electors.
- The estimated number of electors in population.
- The estimated number of voters.
- The percentage of voters among electors in the population and its 95% confidence limits.
- The percentage of voters among the registered electors and its 95% confidence limits.

Preliminary Analysis

Figure 5 summarizes the gender distribution among the voters by Canadian federal election. It shows that in the three most recent federal elections, there are always more female voters participated than male voters. The percentage of male voters decreases from year 2008 to year 2015 (48.2% vs. 47.6%), while the percentage of female voters increases during this period (51.8% vs. 52.4%). The ratio of male voters over female voters decreases (0.93 in year 2008, vs. 0.91 in year 2015).

Figure 6 summarizes the distribution of age groups among the voters for each election. The percentage of middle-aged voters, i.e., the age groups of 35-44 and 45-54 decrease over the years (17.10% in 2008 vs. 14.95% in 2015, 21.47% in 2008 vs. 18.26% in 2015, respectively). Younger voters, that is the age groups of 18-24 and 25-34 participated more in the most recent election (7.90% in 2011 vs. 9.62% in 2015, 13.78% in 2011 vs. 14.18% in 2015, respectively). But the older voters show various trends, the percentages of voters in age groups of 55-64 and 65-74 increase over the years (18.20% in 2008 vs. 19.57% in 2015, 11.49% in 2008 vs. 14.15% in 2015, respectively). The most senior group of age 75 or above shows a decreasing trend (10.06% in 2008 vs. 9.27% in 2015).

2.2 Federal Election Survey

2.2.1 Post-Election Survey of Elector

After the federal general election, the Elections Canada commissions a company to conduct a survey with electors eligible to vote in the election. The purpose is to evaluate the public experience, attitudes, and knowledge about the electoral process, as well as to obtain opinions about electors' experiences during the election (Elections Canada, 2011). The survey questionnaire covers several main areas including elector awareness, registration experience, service experience, etc. Particularly, we are interested in the following two questions included in the survey questionnaire: whether the elector voted in the current election, and what is the main reason for not voting. The survey of electors conducted (Elections Canada, 2011) for the 41^{st} federal election held in year 2011 is summarized in this section.

Survey Respondents

Table 4 presents the socio-demographics of survey respondents. In total, 3,570 Canadians who are eligible to vote in the 41^{st} federal election are included as the final sample. There are 504 youth electors (18-24 years old) and 528 Aboriginal electors in the sample. 49% of the respondents are males and 51% are females. More than one-third of the respondents 25-44 years old (35%) or 45-64 years old (34%). More than one-third (33%) the respondents have university degree.

Voting in the Election

One question in the survey asks "Many people don't or can't vote for a variety of reasons. Did you vote in the May 2^{nd} federal election?". Among all the electors in the sample, a large majority (84%) reported having voted in the 41^{st} general election. Particularly, 69% of youth electors and 67% of Aboriginal electors voted in the election. A positive relationship showed between the elector's age and having reported voting. Electors who stay at home full-time are more likely to report having voted than others. (89% vs. 77-83% of others). Electors who have more than high school education are more likely to report having voted (85-89% vs. 78% with high school or less). And electors with household incomes of at least $40,000 are more likely to report having voted than those with household incomes less than $40,000 (85-90% vs. 79%).

Reason for Not Voting

The 512 respondents from the sample reported did not vote in the 41^{st} federal general election were asked to identify the main reason why they did not vote. The main reasons are grouped into four main categories: everyday life issues, political issues, electoral process issues, and other issues. The majority identified everyday life issues (60%) as the main reason for not voting. 30% of the sample identified political issues as the main reason, while only a small proportion identified the main reason as electoral process issues or other issues (6% and 2%, respectively).

Electors aged 45-64 were the least likely to choose the everyday life issues as the main reason (45% vs. 61-67% of others). They were the most likely to take political issues (46% vs. 25-28% of others) as the main reason. Compared to women, men were less likely to say the everyday life issues (53% vs. 67%) was the main reason why they did not vote, but more likely to say they did not vote because of political issues (37% vs. 23%).

2.2.2 Opinion Polls

Many research companies conduct opinion polling months before the election day until the day before the election day. For the opinion poll, a random sample of eligible electors will be selected across Canada. The main question the respondents will be asked is which party they are most likely to vote for, or have already voted for, in the federal election on election day. The respondents' socio-demographics will also be collected, such as gender, income, education, etc. From the data collected by the polling, the percentage of the votes that each major political party received will be calculated.

Polls for the 42^{nd} Canadian Federal Election

The 42^{nd} election was held on 19 October 2015. The opinion polls were conducted from 02 August 2015 till 18 October 2015, the day before election day. For this election, the Conservatives received around 31.9% of the votes, Liberals received around 39.5%, NDP received around 19.7%, Bloc Québécois received about 4.7%, and the Green party received about 3.4% of the votes.

Table 2 summarizes the results of the opinion polls conducted by different research companies from 17 Oct 2015 to 19 Oct 2015 (Wikipedia contributors, 2019), aimed for the federal election happened on 19 October 2015. The denominator for the percentages in this table is the sample size which is listed in the last column. All these seven opinion polls show the Liberals receives the highest percentage of the votes, which is consistent with the real outcomes of the 42^{nd} Canadian federal election.

Chapter 3

Canadian Community Health Survey

3.1 Data Overview

CCHS is a national survey created together by the Canadian Institute for Health Information, Statistics Canada and Health Canada, to respond to issues and problems associated with the health information system in Canada. The survey began collecting data since 2001 and was repeated every two years until 2005. From 2007, the CCHS data were collected annually (Statistics Canada, 2018). This national survey uses a cross-sectional design and gathers health-related data. Its target population is 12 years of age and over living in the ten provinces and the three territories in Canada. Data are collected directly from the survey respondents. For the cycles of 2001, 2003 and 2005, the sample size is approximately 130,000 respondents; the sample size was changed to 65,000 respondents annually since 2007. The primary use of the CCHS data is for health surveillance and population health research (Statistics Canada, 2018). This survey mainly covers the following areas: diseases and health conditions, lifestyle and social conditions, prevention, and detection of disease. In this section, we will be focusing on the surveys conducted from the year 2007 to 2018.

3.2 Variables Related to Life Satisfaction

Life Satisfaction

Among the variables provide by CCHS, we are mainly interested in the survey question about Life Satisfaction. This survey question asks the respondents how satisfied they are with their lives in general. For the CCHS conducted in year 2007 and 2008, the following five options are provided to the respondents: "very satisfied", "satisfied", "neither satisfied nor dissatisfied", "dissatisfied" and "very dissatisfied". Since year 2009, this survey question asks the respondents to give a score from 0 to 10 to evaluate their satisfaction with life in general. We categorize the scores into the same five categories used for the CCHS before year 2009.

In addition to the five categories and 0-10 scale mentioned above, this question also has the other three options: "don't know", "refusal" and "not stated". All these three options are grouped into one category "NA" in our analysis.

Socio-demographics

There are many socio-demographics variables available in CCHS. Based on the interests of this project, we select the following four variables: gender, age, highest education level, and annual household income.

- Gender
  There are two options for this question: male and female.

- Age
  The survey asks the respondent what his/her age is. The answer to this question is a positive integer. Since only the Canadian citizen who is 18 years old or older has the right to vote, we select the CCHS respondents who are 18 years old or older into our analysis and categorize the age into the following seven groups: 18-24, 25-34, 35-44, 45-54, 55-64, 65-74 and 75+.

- Education Level
  This survey question asks the respondents about their highest level of education. The options to this question are categorized into the following four categories: "no post-secondary degree/certificate/diploma", "Diploma/Certificate", "Bachelor's Degree or above" and "NA". The "NA" category contains "not applicable", "don't know", "refusal" and "not stated".

- Annual Household Income
  This survey question asks the respondents their estimated household income in the past 12 months. The answer is a positive integer. We categorize the annual household income into the following six categories in CAD: <20,000, 20,000-40,000, 40,000-60,000, 60,000-80,000, 80,000-100,000 and >100,000.

3.3 Preliminary Analysis of Selected Variables

3.3.1 Descriptive Analysis

Table 3 and Figure 7 summarize the distribution of life satisfaction among the respondents cross Canada for each CCHS from year 2007 to 2018. The majority (over 87%) of the respondents feel "satisfied or very satisfied" about their lives in general. About 5% to 6%

of the respondents choose the neutral answer, i.e., "neither satisfied nor dissatisfied". And around 3% of the respondents feel "dissatisfied or very dissatisfied" about their lives.

Figure 8 summarizes the gender distribution among the respondents in CCHS. From year 2007 to 2018, there are always more females responded to the survey than males. The distribution of gender among the respondents is pretty stable over the years. The percentages of females respondents are around 55% while the percentages of male respondents are around 45%.

Figure 9 summarizes the distribution of age groups among the respondents in CCHS. Throughout the eight years, there are more senior respondents in CCHS, i.e., from age group 55-64 (17.69% in year 2007 vs. 19.15% in year 2018), 65-74 (12.97% in year 2007 vs. 18.88% in year 2018), and 75 or above (11.62% in year 2007 vs. 13.16% in year 2018). The percentages of younger respondents in CCHS decrease, i.e., from age group 25-34 (15.05% in year 2007 vs. 14.01% in year 2018), 35-44(16.78% in year 2007 vs. 14.15% in year 2018), and 45-54 (17.55% in year 2007 vs. 14.12% in year 2018). The percentage of the youngest age group 18-34 also decrease over the years (8.34% in year 2007 vs. 6.53% in year 2018).

Figure 10 summarizes the distribution of the highest education level among the CCHS respondents. The percentage of people with "Bachelor's degree or above" shows an increasing trend over the years (16.90% in year 2007 vs. 22.83% in year 2018), in the meanwhile the percentage of people with "no post-secondary degree/certificate/diploma" decreases over the eight-year-period (7.35% in year 2007 vs. 4.43% in year 2018).

Figure 11 summarizes the annual household income distribution among the respondents. Note that the option of "NA" has been removed from the survey since year 2011, therefore all the other six groups increase a lot from year 2010 to year 2011. The percentages of lower-income groups, i.e., <20,000 (11.27% in year 2011 vs. 8.56% in year 2018), and 20,000-40, 000 (23.75% in year 2011 vs. 18.23% in year 2018) decrease over the years. The percentage of the highest income group, 100,000 or above shows an increasing trend (21.42% in year 2011 vs. 32.65% in year 2018).

3.3.2 Inferential Analysis

Temporal Trend Analysis

The Cochran-Armitage test is used to assess the statistical significance of temporal trends. We assess the temporal trend in the percentage of each category under the respondents' life satisfaction, gender, age, and education level from the year 2007 to 2018. The percentage of respondents feel satisfied or very satisfied with their lives in general shows a significant decreasing temporal trend ($p<0.0001$). The percentage of respondents feel dissatisfied or very dissatisfied also shows a significant decreasing temporal trend ($p<0.0001$). While the percentage of respondents feel neither satisfied nor dissatisfied shows a significant increasing temporal trend ($p=0.0007$).

The percentage of male respondents shows a significant decreasing temporal trend ($p<0.0001$). The percentages of respondents in the age group 25-34, 35-44, and 45-54 show significant decreasing temporal trends ($p<0.0001$, $p<0.0001$, and $p<0.0001$, respectively). The percentages of respondents in the age group 55-64, 65-74 and 75 or above show significant increasing temporal trends ($p<0.0001$, $p<0.0001$, and $p<0.0001$, respectively). The percentage of respondents in the age group 18-24 does not show any significant temporal trend ($p=0.4773$). The percentages of respondents that do not have a post-secondary degree/certification/diploma, and that have certificate/diploma show significant decreasing temporal trends ($p<0.0001$ and $p=0.0004$, respectively). While the percentage of respondents have bachelor or above degree shows a significant increasing temporal trend ($p<0.0001$).

Logistic Regression Analysis

To explore how the socio-demographics affect life satisfaction among Canadians, we conduct the logistic regression analysis. The socio-demographics from CCHS used in the regression as the covariates are: gender (ref: female), age group (ref: 18-24), the highest education level (ref: no post-secondary degree), and annual household income (ref: <20,000 CAD).

The response variable is a binary variable W_i, it equals to 1 if the i_{th} subject feels satisfied or very satisfied about life in a certain year, otherwise it equals to 0. Let G_i denote a 5-dimensional vector of observed covariates with the first element being 1. The other four elements of G_i are the four socio-demographics mentioned above for the i^{th} subject in the same year. We assume W_i follows a Bernoulli(η_i) distribution where η_i denotes the probability that that the i^{th} subject feels satisfied or very satisfied about life , given the observed covariates G_i, i.e. $\eta_i = Pr(W_i = 1|G_i)$. The regression model is as follows.

$$\log(\frac{\eta_i}{1-\eta_i}) = G_i'\nu$$

We use the data of CCHS in year 2011 to fit the model and the analysis results are summarized in Table 5. All the survey respondents cross the whole nation are included in the logistic regression model. Compare to female respondents, male respondents are less likely to feel satisfied. In terms of age, compare to the youngest group aged 18-24, other people are less likely to feel satisfied, and the level of this likelihood increases as the age increases. Respondents who have a certificate, diploma, bachelor's degree or above are more likely to feel satisfied compared to those who do not have any post-secondary degree. Compare to the lowest annual household income group <20,000 CAD, the other groups of respondents are more likely to feel satisfied, and the level of likelihood increases as the income increases.

3.4 Comparing CCHS Respondents with Election Voters

Figure 13 compares the comparison of the gender distribution between CCHS respondents and Canadian federal election voters. The y-axis showed the ratio of males over females among survey respondents or election voters. For both CCHS and federal election, this ratio is always smaller than one, i.e., there are always more females participated in the survey or election. The differences between the number of males and females are more significant among the survey respondents than those among the election voters. This indicates that the CCHS population is not the same as the election voter population.

Figure 14 compares the distribution of age group between CCHS respondents and federal election voters. We further group the survey respondents and election voters into three categories: young adults (18-34), middle-aged adults (35-64), and older adults (65+). For middle-aged adults group and older adults group, the survey respondents and election voters show similar proportions. Both respondents and voters show a decreasing trend in middle-aged adults group and an increasing trend in older adults group.

We further explore how life satisfaction changes over the years in which the federal election was held. In Figure 12, the y-axis is the percentage of respondents in Canada feel satisfied with their lives. We first fit several curves and line to model the change for the whole period from the year 2007 to 2018. The green line is the fitted simple linear regression line, and the red dashed-curve is the fitted linear regression model with a quadratic term. As the figure shows, the relation between life satisfaction and time is not linear but quadratic. Since there are three federal elections held within this period, the year 2008, 2011, 2015 and 2019, we further model the changing pattern of three sub-periods, year 2008 to 2011, year 2011 to 2015 and year 2015-2018, respectively. The red curves are the fitted linear regression curve with a quadratic term with all the four time-points within each period. The blue lines are the fitted simple linear regression lines for the most recent three time-points within each period. These lines show that within these two periods life satisfaction depends on the year linearly.

Chapter 4

Jointly Modeling Election Outcomes and Longitudinal Measures of Life Satisfaction

In this section, we introduce joint models for longitudinal measures and a binary indicator. Specifically, the joint models approach consists of two models: the first model is a linear mixed effects model for the continuous longitudinal measures of life satisfaction obtained from CCHS, the second model is a generalized linear model for a binary indicator of election outcomes. These two models are linked to each other through a shared random effect.

4.1 Formulation and Modeling

4.1.1 Notation

For riding $i = 1, 2, ..., n$ and discrete time point $j = 1, 2, ..., k$, we define the following:

- Define the primary response variable R_i as a binary indicator of changes to election results in the i^{th} riding. For the i^{th} riding, if the elected political party in the current election changed from the previous election then R_i is 1, otherwise R_i is 0.

- Let the discrete time point j represent the $(j-1)^{th}$ year from the previous election year, e.g. if the previous election is held in year 2008, then $j = 1$ represents the 0^{th} year from year 2008, which is also year 2008, $j = 2$ represents the 1^{st} year from year 2008, which is year 2009, etc. The maximum value that j could take on is k, $j = k$ represents the current election year, which is the $(k-1)^{th}$ year from the previous election year. That is, if the current election is held in year 2011 then $k = 4$, which means year 2011 is the 3^{rd} year from the previous election year.

- Let Y_{ij} denote in the i^{th} riding at time point j, the logit transformed proportion of respondents that feel very satisfied or satisfied with their lives in general, derived from the CCHS data. Define $Y_i = (Y_{i1},Y_{i2},...,Y_{ik})'$ as the k dimensional vector of continuous longitudinal measures of life satisfaction for the i^{th} riding.

4.1.2 Joint Models

Modeling CCHS Based Longitudinal Life Satisfaction

The first model of the continuous longitudinal measurements Y_i can be written as a linear mixed effects model (Kleinbaum et al., 2014):

$$Y_i = X_i\alpha + Z_i b_i + \varepsilon_i, \tag{4.1}$$

where X_i is an observed covariates matrix with k rows, and α is a vector of the fixed effect parameters. For X_i, the 1^{st} column is all $1'$s, the other columns contain the time factor, and the longitudinal measures of socio-demographics obtained from CCHS for the i^{th} riding at each discrete time point, including gender, age, the highest education level, and the annual household income.

We define Z_i as the observed covariates matrix with k rows, with the first column being all $1'$s and the second column being the discrete time point j. Matrix Z_i is a subset of matrix X_i. b_i is a vector of the random effects for the i^{th} riding, accounts for the effect of the natural heterogeneity among ridings. ε_i is a k dimensional vector of random errors for the i^{th} riding. We assume that

- $b_i \overset{iid}{\sim} N(0, D)$,
- $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2 I_k)$,
- b_i and ε_i are mutually independent.

Modeling Election Outcomes

The second model for the primary response variable R_i the binary indicator of the election outcome, is assumed to be a generalized linear model with a logit link function (Agresti, A., 2014).

$$log(\frac{\pi_i}{1-\pi_i}) = U_i'\beta + b_{i1}\gamma, \tag{4.2}$$

where U_i is a component vector of the corresponding matrix X_i in life satisfaction model. U_i is an observed covariates with the 1^{st} element being 1. The other columns contain the socio-demographics covariates of voters for the i^{th} riding in the current election year. β is a vector of the unknown parameters. b_{i1} is a component vector of corresponding vector b_i in life satisfaction model. b_{i1} summarizes the input of life satisfaction on the election outcomes. γ is a vector of unknown parameters.

We assume R_i follows a Bernoulli(π_i) distribution where π_i denotes the probability that that the incumbent political party in the i^{th} riding was not elected in the current election, given the observed covariates U_i and b_{i1}, i.e., $\pi_i = Pr(R_i = 1|U_i, b_{i1})$.

Joint Models

We assume R_i and Y_i are independent conditional on the observed covariates and random effects, for $i = 1, ..., n$. The second model of election outcomes is linked to the first model for longitudinal CCHS data through their common random effect b_{i1}. The parameter γ measures the strength of the association between these two models. With the joint models, the longitudinal measures of life satisfaction from CCHS are linked to the election outcomes through the common random effect, and we are able to predict the election outcomes with using the information from both CCHS and election.

We consider two different estimation approaches for the joint models: one is a two-stage approach, and the other one is a maximum-likelihood based approach. We start with the two-stage estimation approach in Section 4.2, in which the estimations of the parameters are computed separately for the longitudinal measures model and the election outcomes model. In Section 4.3, we discuss the maximum-likelihood estimation approach, specifically, we apply the Monte Carlo Expectation Maximization (MCEM) algorithm to estimate the unknown parameters. Both estimates of the two-stage approach and the MCEM approach are presented in Section 4.4.

4.2 Two-Stage Estimation

The joint models could be fitted with two stages, and the parameter estimation is conducted separately for the life satisfaction model and the election outcomes model, as follows (Horrocks & Heuvel, 2009). At the first stage, the estimates of parameters and the predicted random effects are obtained from the life satisfaction model without consideration of the election outcomes. At the second stage, the predicted random slope obtained in the first stage is used as the known predictor in the election outcomes model.

Stage I

At the first stage, we fit a linear mixed effects model to the longitudinal measures of life satisfaction and compute the estimated fixed effect parameter $\hat{\alpha}$, the estimated variance of random error $\hat{\sigma}^2$, the predicted random effects $\hat{b}_i$, and the estimated variance-covariance matrix of random effects $\hat{D}$. Particularly, we use SAS 9.4 (SAS Institute Inc., 2014) Proc

Mixed procedure to fit the linear mixed model and obtain the estimates and predictions, as follows.

We use the restricted/residual maximum likelihood (REML) approach to obtain the estimates of random error σ^2 and the variance-covariance matrix of random effects D. The estimates are obtained by maximizing the following log-likelihood function (SAS Institute Inc. 2008.).

$$l(D, \sigma^2) = -\frac{1}{2}\log|V_i| - \frac{1}{2}\log|X_i{}'V_i{}^{-1}X_i| - \frac{1}{2}r_i{}'V_i{}^{-1}r_i - \frac{n-p}{2}\log(2\pi),$$

where V_i denotes the variance of Y_i, and

$$V_i = Z_i D Z_i{}' + \sigma^2 I_k,$$
$$r_i = Y_i - X_i(X_i{}'V_i{}^{-1}X_i)^{-1}X_i{}'V_i{}^{-1}Y_i.$$

With the estimates of σ^2 and D, we estimate the unknown parameters α and predict the random effects b_i as follows,

$$\hat{\alpha} = (X_i{}'\hat{V_i}^{-1}X_i)^{-1}X_i{}'\hat{V_i}^{-1}Y_i,$$
$$\hat{b_i} = \hat{D}Z_i{}'\hat{V_i}^{-1}(Y_i - X_i\hat{\alpha}).$$

Stage II

At the second stage, we fit a logistic regression model to the election outcomes model and use the best linear unbiased predictors (BLUPs) of the random slope obtained from the first stage as the known covariates. Specifically, we use SAS 9.4 (SAS Institute Inc., 2014) Proc Logistic procedure to fit the election outcomes model and compute the estimated parameters $\hat{\beta}$ and $\hat{\gamma}$.

4.3 Maximum Likelihood Estimation

In addition to the two-stage approach, we consider using the maximum likelihood approach to estimate the parameters of the joint models. Section 4.3.1 introduces the likelihood function for the joint models based on the full data given the random effects. In Section 4.3.2, the MCEM algorithm is applied to estimate the parameters.

4.3.1 Likelihood Function

The full data consist of the longitudinal measures Y_i, the binary indicator R_i, and the random effects b_i which links the two measurements together. While in the full data, only Y_i and R_i are observed, i.e., they are the observed data, the random effects b_i is unobserved. Let $\theta = (\alpha, \beta, \gamma, D, \sigma^2)$ represent all the parameters need to be estimated.

The likelihood function based on the observed data is:

$$L_{obs}(\theta) = \prod_{i=1}^{n} \int_{-\infty}^{\infty} p(Y_i|X_i, b_i; \theta) p(R_i|U_i, b_{i1}; \theta) p(b_i|\theta) db_i. \tag{4.3}$$

It is difficult to directly maximize the above likelihood function given that it contains the complex integral computation. Therefore, we employ the EM algorithm (Dumpster et al., 1977) to compute the MLE of θ as follows. We view the full data are the observed data in combination with the random effects b_i. It yields the following likelihood function of θ based on the full data:

$$L_F(\theta) = \prod_{i=1}^{n} p(Y_i|X_i, b_i; \alpha, \sigma^2) p(R_i|U_i, b_{i1}; \beta, \gamma) p(b_i; D), \tag{4.4}$$

where

$$\begin{aligned}
p(Y_i|X_i, b_i; \alpha, \sigma^2) &= (\frac{1}{\sqrt{2\pi}})^k |\sigma^2 I_k|^{-1/2} \exp[-\frac{1}{2}(Y_i - X_i\alpha - Z_i b_i)' (\sigma^2 I_k)^{-1} (Y_i - X_i\alpha - Z_i b_i)], \\
p(R_i|U_i, b_{i1}; \beta, \gamma) &= {\pi_i}^{R_i} (1 - \pi_i)^{R_i} \text{ with } \mathrm{logit}(\pi_i) = U_i'\beta + b_{i1}\gamma, \\
p(b_i; D) &= \frac{1}{2\pi} |D|^{-\frac{1}{2}} \exp(-\frac{1}{2} {b_i}^T D^{-1} b_i).
\end{aligned}$$

Therefore the log-likelihood function based on the full data given the random effects b_i is:

$$\begin{aligned}
l_F(\theta) &= \log(L_F(\theta)) \\
&= \sum_{i=1}^{n} [\log p(Y_i|X_i, b_i; \theta) + \log p(R_i|U_i, b_{i1}; \theta) + logp(b_i; \theta)],
\end{aligned} \tag{4.5}$$

where

$$\begin{aligned}
l_{F1}(\alpha, \sigma^2) &= \sum_{i=1}^{n} \log p(Y_i|X_i, b_i; \alpha, \sigma^2), \\
l_{F2}(\beta, \gamma) &= \sum_{i=1}^{n} \log p(R_i|U_i, b_{i1}; \beta, \gamma), \\
l_{F3}(D) &= \sum_{i=1}^{n} \log p(b_i; D).
\end{aligned}$$

The EM algorithm contains two steps: the expectation (E) step and the maximization (M) step. At E step, the expectation of the log-likelihood function is calculated using the current estimates of the parameters. Then at the M step, the estimates of the parameters are updated by maximizing the expectation function computed in the E step. These two steps are repeated until the estimates converge. However, in our joint models, the expectation of the log-likelihood function is not straightforward to be computed. Therefore, we introduce another version of the EM algorithm, the MCEM algorithm (McCulloch, 1997; Levine & Casella 2001). The difference between these two algorithms is in the E step. The details are presented in the next section.

4.3.2 Monte Carlo Expectation Maximization Algorithm

MC-E Step

In the MC-E step, we first calculate the expected log-likelihood function based on the full data given the conditional distribution of random effects, provided the current value of θ, i.e., $\theta^{(c)}$, c represents the c^{th} iteration.

$$\begin{aligned} Q(\theta|\theta^{(c)}) &= E[l_F(\theta|\underline{b})|\underline{Y}, \underline{R}, \theta^{(c)}] \\ &= E[l_{F1}(\alpha, \sigma^2|\underline{b})|\underline{Y}, \underline{R}, \alpha^{(c)}, \sigma^{2^{(c)}}] \\ &+ E[l_{F2}(\beta, \gamma|\underline{b_1})|\underline{Y}, \underline{R}, \beta^{(c)}, \gamma^{(c)}] \\ &+ E[l_{F3}(D|\underline{b})|\underline{Y}, \underline{R}, D^{(c)}], \end{aligned} \tag{4.6}$$

where $\underline{b}$ is a vector of random effects b_i, $\underline{b_1}$ is a vector of random slope b_{i1}, $\underline{Y}$ is a vector of the life satisfaction measures Y_i, and $\underline{R}$ is a vector of the indicators for election outcomes R_i, $i = 1, ..., n$.

On each iteration, we draw $\mathbf{b}^{(1)}, \mathbf{b}^{(2)}, ..., \mathbf{b}^{(M)}$ from $p(b_i|Y_i, R_i; \alpha^{(c)}, \sigma^{2^{(c)}}, D^{(c)})$, where $\mathbf{b} = (b_1, b_2, ..., b_n)'$. We approximate the above Q-function as:

$$\begin{aligned} Q(\theta|\theta^{(c)}) &\approx \frac{1}{M}\sum_{m=1}^{M} l_F(\theta|b^{(m)}|\underline{Y}, \underline{R}, \theta^{(c)}) \\ &= \frac{1}{M}\sum_{m=1}^{M}[l_{F1}(\alpha, \sigma^2|b^{(m)}) + l_{F2}(\beta, \gamma|b^{(m)}) + l_{F3}(D|b^{(m)})], \end{aligned} \tag{4.7}$$

where

$$\begin{aligned} l_{F1}(\alpha, \sigma^2|b^{(m)}) &= -\frac{nk}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{n}[(Y_i - X_i\alpha - Z_ib_i)^T(\sigma^2 I_k)^{-1}(Y_i - X_i\alpha - Z_ib_i)], \\ l_{F2}(\beta, \gamma|b^{(m)}) &= \sum_{i=1}^{n}[R_i(U_i{}'\beta + b_{i1}\gamma) - \log(1 + e^{U_i{}'\beta + b_{i1}\gamma})], \\ l_{F3}(D|b^{(m)}) &= -n\log(2\pi) - \frac{n}{2}\log|D| - \frac{1}{2}\sum_{i=1}^{n} b_i{}'D^{-1}b_i. \end{aligned}$$

MC-M Step

In the MC-M step, we maximize $Q(\theta|\theta^{(c)})$ to update the estimates of the current parameters (R Core Team, 2017).

$$\theta^{(c+1)} = \underset{\theta}{argmax} Q(\theta|\theta^{(c)}). \tag{4.8}$$

The E step and M step are repeated until convergence of $\theta^{(c)}$, c=1,2,..., i.e.

$$\lim_{c\to\infty} \theta^{(c)} = \hat{\theta}.$$

In order to effectively apply this MCEM Algorithm, we make use of the estimates computed from the two-stage estimation approach as follows.

(1) Estimate the conditional distribution of random effects b_i given the observed longitudinal measures in CCHS, i.e., $p(b_i|Y_i, X_i; \hat{\alpha}, \hat{D}, \hat{\sigma}^2)$.

We derive the following joint distribution of random effects b_i and observed data $Y_i - X_i\alpha$

$$\begin{pmatrix} b_i \\ Y_i - X_i\alpha \end{pmatrix} \sim MN_{2+k} \left(\mathbf{0}, \begin{bmatrix} D & D{Z_i}' \\ Z_iD' & Z_iD{Z_i}' \end{bmatrix} \right). \tag{4.9}$$

Then the conditional distribution of random effects is estimated as

$$(b_i|Y_i - X_i\alpha) \sim MN_2({\mu_i}^*, {\epsilon_i}^*) \text{ where}$$

$${\mu_i}^* = \hat{D}{Z_i}'(Z_i\hat{D}{Z_i}' + \hat{\sigma}^2 I_k)^{-1}(Y_i - X_i\hat{\alpha}),$$

$${\epsilon_i}^* = \hat{D} - \hat{D}{Z_i}'(Z_i\hat{D}{Z_i}' + \hat{\sigma}^2 I_k)^{-1} Z_i\hat{D}'.$$

(2) Use the estimates as the initial values.

In the first iteration, we need to set the initial values $\theta^{(0)}$ for the parameters. Usually the initial values are provided based on the best guess. Since we already obtain the estimated parameters from the two-stage estimation approach, we use these estimates as the initial values to be used in the first iteration of the MCEM approach.

4.4 Results and Discussion

We assume the current election year is 2011, therefore the primary endpoint R_i is derived from the Canadian federal election outcomes in the year 2011 compare to the outcomes in the previous election in the year 2008. For riding i, if the winning party for election in the year 2011 is different from the outcomes of the election in the year 2008, then R_i is 1, otherwise R_i is 0. The 2011 federal election related data and the CCHS conducted between the year 2008 and 2011 are used to fit the joint models.

We first fit the life satisfaction model (4.1) with assuming the two random effects, the random intercept and random slope are dependent on each other. In the results, the estimated covariance of these two random effects is very small and not statistically significant ($p>0.05$). Therefore we assume the two random effects are independent and fit the joint models as follows.

4.4.1 Approximations to Predictor Values

In the election outcomes model (4.2), two observed socio-demographics predictors included are the percentage of male voters and an age-related factor of voters in the i^{th} riding for the current election year. These two predictors are not available directly in the data because the voter information is only provided on the province level, not on the riding level. Therefore We propose the following four methods to estimate these two predictors.

Method 1: Using Only Election Information.

We assume all the ridings located in the same province have the same distribution of gender and age among the voters as the province has. That is, the percentage of male voters of the i^{th} riding in the current election year is estimated by the percentage of male voters of the province where the i^{th} riding belongs to in the same year. Moreover, the age-related predictor of the i^{th} riding in the current election year is estimated by the age distribution of the province where the i^{th} riding belongs to in the same year.

Method 2: Using Only CCHS Information

We assume the voters in each riding have the same gender and age distribution as the respondents to the CCHS have in the same riding. That is, the percentage of male voters of the i^{th} riding in the current election year is estimated by the percentage of male respondents of CCHS in the same riding in the same year. Also, the age-related predictor of the i^{th} riding in the current election year is estimated by the age distribution among the respondents of CCHS for the same riding in the same year.

Method 3: Using both Election and CCHS Information, Separately.

Although the gender and age information of the voters are only available on the province level, this part of the information is available on the riding level among the respondents in CCHS. We assume the election voters and the survey respondents have the following relation. Take the percentage of male voters of the i^{th} riding in the current election year as an example, assume the i^{th} riding belongs to the l^{th} province, then we have the following equation for the current election year:

$$\frac{\text{Province l: \% of male voters}}{\text{Province l: \% of male respondents}} = \frac{\text{Riding i: \% of male voters}}{\text{Riding i: \% of male respondents}}$$

In the above equation, only the percentage of male voters of the i^{th} riding in the current election year is unknown. Therefore we use the other three available percentages to estimate this unknown percentage. Similarly, we estimate the age-related predictor among the voters for the i^{th} riding in the current election year.

Method 4: Using both Election and CCHS Information, Jointly.

(4) In the above estimation method, we estimate the gender and age groups among the voters on riding level separately. Here we consider another estimation method which estimates the combination of gender and age groups among the voters on riding level. Take the percentage of young male voters of the i^{th} riding in the current election year as an example, assume the i^{th} riding belongs to the l^{th} province, then we have the following equation for the current election year:

$$\frac{\text{Province l: \% of young adults male voters}}{\text{Province l: \% of young adults male respondents}} = \frac{\text{Riding i: \% of young adults male voters}}{\text{Riding i: \% of young adults male respondents}}$$

In the above equation, only the percentage of young adults male voters of the i^{th} riding in the current election year is unknown. Therefore we use the other three available percentages to estimate this unknown percentage. Similarly, we estimate the percentage of voters for other combinations of gender and age groups for the i^{th} riding in the current election year. Then the estimated marginal distribution of gender and age group among the voters on riding level is calculated.

4.4.2 Model Selection

4.4.2.1 Time Variable

To address the changing in life satisfaction and the changing in different age groups among Canadians over the years, we consider the following four different model settings to fit the longitudinal life satisfaction model (4.1).

1. Use the four years of CCHS data between the two federal elections to fit the model, i.e., from year 2008 to 2011.

2. Based on model setting 1, include an interaction effect between the age-related predictor and the linear term of year.

3. Use the most recent three years of CCHS data between the two federal elections to fit the model, i.e., from year 2009 to 2011. Remove the quadratic term of year from the predictors.

4. Based on model setting 3, include an interaction effect between the age-related predictor and the linear term of year.

We fit the life satisfaction model with the above four settings and use the percentage of survey respondents aged 18-34 as the age-related predictor. We compare the MCEM estimation results between model setting 1 and 3, i.e., the four-year model setting and the three-year model setting. Except for the time-related factors, these two settings give similar estimates. In the four-year model setting, the estimated parameter for the linear term of year is negative and significant (-0.8824, $p<.0001$), and the estimated parameter for the quadratic term of year is positive and significant (0.7589, $p<.0001$). In the three-year model setting, the estimated parameter for the linear term of year is positive and significant (0.1865, $p=0.0003$). We consider the three-year model can better summarize the changing in life satisfaction among the Canadians over the recent three years and its effect on the federal election outcomes. The Canadian federal election in year 2008 was held in October, whereas the CCHS in year 2008 was conducted throughout the whole year. So the information of the life satisfaction status provided in CCHS 2008 is more related to the federal election in the same year other than the current election in year 2011. The life satisfaction status among Canadians in the most recent three years is more likely to affect their voting behavior during the current federal election. As Figure 12 shows, the percentage of Canadians feel satisfied has an increasing linear trend from the year 2009 to 2011.

We then compare the MCEM estimation results between model setting 3 and 4, i.e., the three-year model without an interaction effect and the same setting with an interaction effect. These two settings produce similar estimation results. The estimated interaction effect between the percentage of survey respondents aged 18-34 and the linear term of year is, however, not significant ($p=0.6693$). Therefore, we choose to use the setting 3 for life satisfaction model, i.e., a three-year longitudinal model without an interaction effect.

4.2.2.2 Age-related Predictor

In both life satisfaction model and election outcomes model, we include an age-related factor as one predictor. For this factor, we consider the following four options.

1. The percentage of young adults (age 18 - 34).

2. The percentage of middle-aged adults (age 35 - 64).

3. The percentage of older adults (age 65 or above).

4. The ratio of young over middle-aged adults.

We fit the joint models with each of the above four age-related predictors. Specifically, as discussed before, we use the three-year model setting without an interaction effect for the life satisfaction model, and use four different predictors estimation methods for the election outcomes model.

The MCEM estimation results of the election outcomes model are compared. For each of the four predictors estimation methods, the fixed effect of the young adults group is positive and significant ($p<0.05$). It indicates that young adults are more likely to vote for a new political party. The fixed effect middle-aged adults group is only significant for predictors estimation Method 1 and 2. With Method 1, only election information, the negative fixed effect (-0.5196, $p=0.0026$) indicates that middle-aged adults are less likely to vote for a new party if the predictors are estimated only with voter information. With Method 2, only CCHS information, the positive estimated effect (0.3030, $p=0.0185$) indicates that middle-aged adults are more likely to vote for another party if the predictors are estimated with survey respondents information only. For each of the four predictors estimation methods, the older adults group shows a negative significant fixed effect ($p<0.05$). It indicates that older adults tend to vote for the same political party. With the ratio of young over middle-aged adults as the age-related predictors, its fixed effect is only significant with predictor estimation Method 1 (55.7767, $p<0.0001$) and marginally significant with estimation Method 3 (2.0941, $p=0.0518$). The results indicate the ridings that have more young adults than middle-aged adults are tend to vote for a new political party, if the predictors are estimated only with voter information, or estimated with voter and survey respondents information separately.

Take the above analysis results into consideration, we choose to use the young adults group as the age-related predictor in the joint models. We also consider using the ratio of young over middle-aged adults as another option for the age-related predictor, as we would like to examine if the relative difference between the number of young adults and middle-aged adults will affect the election outcomes.

4.4.3 Analysis Results

Based on the above discussion, we choose to use the model setting 3 to fit the life satisfaction model, i.e., uses the most recent three years of CCHS data without including the interaction effect between age-related predictor and the linear term of year. For the election outcomes model, we use four different estimation methods to estimate the predictors. In addition, we choose to use the percentage of young adults and the ratio of young over middle-aged adults as the two options for age-related predictors in the joint models. The results of both the two-stage estimation approach and MCEM approach are summarized and discussed as follows.

4.4.3.1 With Young Adults Group as Age-related Predictor

With taking the percentage of young adults group as the age-related predictor, we first estimate the parameters for the joint models using the two-stage estimation approach. The estimation results for life satisfaction model are summarized in Table 6. The percentage of respondents feel satisfied with their lives shows a positive linear dependence of year (0.1867, p=0.0014). It also positively depends on the percentage of respondents who are young adults (0.03928, p=0.0064) and with certificate or diploma as the highest level of education (0.03588, p=0.0111). However, the respondents that have a lower annual income (<40,000 CAD) are less likely to feel satisfied with their lives (-0.05400, p=0.0004). The estimated variance of the random slope is not significant (p>0.05).

The estimation results of election outcomes model are summarized in Table 7, Table 8, Table 9, and Table 10 for each predictor estimation method, respectively. Except for the predictors estimation Method 1, only election information, the fixed effect of changing in life satisfaction is positive and statistically significant (p<0.05) in all the other three estimation methods. It indicates that ridings with a higher random slope for life satisfaction tend to have a higher probability of voting for a new political party. For each of the four methods, the fixed effect of male voters is negative and statistically significant (p<0.05). It indicates that male voters are less likely to vote for a new political party. Except for the predictors estimation Method 2, only CCHS information, the fixed effect of the young adults group is positive and statistically significant (p<0.05) in all the other three estimation methods. It indicates that young voters are more likely to vote for another political party.

In addition to the two-stage approach, we estimate the parameters by applying the MCEM algorithm. The results are summarized in Table 11, Table 12, Table 13, and Table 14 for each predictor estimation method, respectively. The MCEM approach produces similar results compared to the two-stage estimation. Two major differences appear in all the four predictors estimation methods. The first one is the fixed effect of changing in life satisfaction no longer statistically significant (p>0.05) with the MCEM estimation approach. The second one is the estimated variance of the random slope is statistically significant (p<0.0001).

4.4.3.2 With Ratio of Young over Middle-aged Adults as Age-related Predictor

With taking the ratio of young adults over middle-aged adults as the age-related predictor, we first estimate the parameters for the joint models using the two-stage approach. The estimation results of life satisfaction model are summarized in Table 15. The percentage of respondents feel satisfied with their lives shows a positive linear dependence of year (0.1958, p=0.0008). It also positively depends on the percentage of respondents with a certificate or diploma as the highest education level (0.03396, p=0.0170). However, the respondents that have a lower annual income (<40,000 CAD) are less likely to feel satisfied with their lives (-0.05860, p=0.0001). The fixed effect of the ratio of young over middle-aged adults is not significant ($p>0.05$). The estimated variance of the random slope is not significant ($p>0.05$).

The estimation results of election outcomes model are summarized in Table 16, Table 17, Table 18, and Table 19 for each predictor estimation method, respectively. For each of the four predictors estimation methods, the fixed effect of male voters is negative and statistically significant ($p<0.05$). It indicates that males are less likely to vote for a new political party. Moreover, the fixed effect of the ration of young over middle-aged adults is statistically significant and positive (51.1737, $p<0.0001$) for predictors estimation Method 1, only election information, but not significant ($p>0.05$) with the other three estimation methods. It indicates that the ridings have more young adults voters than middle-aged voters, then those ridings are more likely to vote for another political party. Except for predictors estimation Method 1, election information only, the fixed effect of changing in life satisfaction is positive and statistically significant in all the other three estimation methods. It indicates that ridings with a higher random slope for life satisfaction tend to have a higher probability of voting for a new political party.

The parameters are also estimated by applying the MCEM algorithm, and the results are summarized in Table 20, Table 21, Table 22, and Table 23 for each predictor estimation method. The MCEM estimation produces similar results compared to the two-stage estimation. Two major differences appear in all the four predictor estimation methods. One is the fixed effect of changing in life satisfaction no longer statistically significant with MCEM approach. The other one is that the variance of the random slope is now statistically significant ($p<0.0001$).

Chapter 5

Examples of Application

5.1 Analysis with Different Primary Response Variable

In Chapter 4, the primary response variable R_i is defined as the indicator of changing in the election outcomes for riding i. If the winning party in riding i for the current election is different from the previous election then R_i is 1, otherwise R_i is 0. The primary response variable could be defined in many other ways, for example, it could be defined as the indicator of the winning party in one riding changes from other political parties to a specific party of interest.

Here, we define a new primary response variable ${R_i}^*$ as an indicator of changing to NDP from other political parties in the current federal election for riding i, compared to the previous election. That is, for riding i, if the winning political party in the current federal election changes from other parties to NDP then ${R_i}^*$ is 1, otherwise ${R_i}^*$ is 0. If for riding i, NDP is the winning party for both the current and the previous elections, then ${R_i}^*$ is 0.

Given that the two-stage approach and MCEM approach give similar estimates, we fit the joint models and obtain the two-stage estimation results. For the age-related predictor used in the joint models, we consider the percentage of young adults group and the ratio of young over middle-aged adults. The life satisfaction model is fitted first with the data of CCHS conducted between 2009 and 2011. The estimation results are summarized in Table 6 and Table 15 for young adults group and for the ratio of young over middle-aged group, respectively. This part of results has been discussed in Chapter 4. The election outcomes model is then fitted for the newly defined primary response variable ${R_i}^*$ with four different predictors estimation methods.

With Young Adults Group as Age-related Predictor

With the young adults age group as one predictor, the two-stage estimation results are summarized in Table 24, Table 25, Table 26, and Table 27 for the four different methods, respectively. For each of the four estimation methods, the fixed effect of changing in life satisfaction is positive and statistically significant ($p<0.05$). It indicates that the ridings with a higher random slope for life satisfaction tend to have a higher probability of changing to vote for NDP. The fixed effect of male voters is negative and statistically significant (-3.5290, $p<0.0001$) with Method 1, election information only, but is not significant with the other three estimation methods. It indicates that males are less likely to change to vote for NDP. Except for Method 2, the fixed effect of the young adults group is positive and statistically significant ($p<0.05$) with all the other three estimation methods. It indicates that young voters are more likely to change to vote for NDP.

With Ratio of Young over Middle-aged Adults as Age-related Predictor

With the ratio of young over middle-aged adults as the age-related predictor, the two-stage estimation results are summarized in Table 28, Table 29, Table 30, and Table 31 for the four predictor estimation methods, respectively. With each of the four estimation methods, the fixed effect of changing in life satisfaction is positive and statistically significant ($p<0.05$). It indicates that the ridings with a higher random slope for life satisfaction tend to have a higher probability of changing to vote for NDP. The fixed effect of male voters is negative and statistically significant ($p<0.05$) with estimation Method 1 and 3, i.e., only election information or using election and CCHS information separately. It indicates that males are less likely to change to vote for NDP. Except for Method 3, using election and CCHS information separately, the fixed effect of the ratio of young over middle-aged adults is positive and statistically significant ($p<0.05$) with all the other three estimation methods. It indicates that if the ridings have more young voters compare to the number of middle-aged voters, then the ridings are more likely to change to vote for NDP.

5.2 Prediction for Canadian Federal Election 2019

The upcoming Canadian federal election will be held on October 2019. With the outcomes of the previous election in the year 2015 and CCHS conducted between the year 2016 and 2018, we predict the upcoming election outcomes by using the two-stage estimation results discussed in Chapter 4.

We first fit the life satisfaction model (4.1) with the most recent three-year of CCHS, i.e., from the year 2016 to 2018. With the fitted model, the predicted random slope for each riding is obtained. To estimate the gender predictor and age-related predictor used in the election outcomes model, we consider the same four methods discussed in Chapter 4. For Method 1, only election information, we use the voter information of the Canadian federal election held in year 2015. We assume the voter information in election 2019 is the same as the previous one in year 2015 since the voter information of election 2019 is not available. For Method 2, we assume the voter information in election 2019 is the same as the respondents' information of the most recent CCHS conducted in the year 2018. For Method

3, we use the voter information of election 2015 and CCHS conducted in the year 2018 to estimate the two predictors separately. And for Method 4, we use the voter information of election 2015 and CCHS conducted in the year 2018 to estimate the two predictors jointly. With the estimated predictors and the two-stage estimation discussed in Chapter 4, we predict the probability of voting for a new political party for each riding in the federal election 2019. For the age factor, we choose to use the young adults group. The prediction results are summarized in Table 32. For each riding, if the predicted probability of voting for a new party is larger than 0.5, then we classify the election result in 2019 for that riding will be changing from the previous election. Since there is no information collected for the three territories in CCHS 2016-2018, we only predict the election outcomes for the other ten provinces, which have 335 ridings in total. Take the Liberals as an example, for the federal election in 2015 the Liberals won 181 ridings. With the only election information estimation method, we predict that out of the 181 seats, 62 of them will change to other political parties from the Liberals in the federal election 2019. With this estimation method, we predict that there will be 109 ridings out of 335 ridings change from the previous elected political parties to other parties in 2019.

To form a majority government, it requires at least 170 seats out of 338 seats. In the federal election 2015, the Liberals won the election in all the three territories. We assume that the election outcomes in these three territories remains the same in the federal election 2019. Therefore, in order to form a majority government, the Liberals need to win 167 seats out of 335 seats. If we choose the estimation method of combining election and CCHS information, we predict 54 seats will be changing from the Liberals to other parties, and for the other four categories of political parties, 57 seats will be changing to other parties. Therefore, the Liberals will be left with 127 seats, and it needs at least 40 out of the 57 seats that changed from other political parties. On the other hand, the Conservatives will lose 29 out of 99 seats to other political parties, and the other four categories of political parties will lose 82 seats to other parties. Therefore, the Conservatives will be left with 70 seats, and it needs at least 100 seats that change from other parties in order to form a majority government. The Liberals appear to be more likely to form a majority government after comparing these two scenarios. The other three estimation methods give similar results.

Chapter 6

Final Remarks

6.1 Conclusion

In this project, we explore the effect of life satisfaction status among Canadians on Canadian federal election outcomes. We propose a joint modeling approach to link life satisfaction with the Canadian federal election outcomes. The joint models are adjusted for important socio-demographics of Canadians. The proposed joint models could be used as predictive models for federal election forecasting.

In Chapter 2, we present and summarize the available data related to Canadian federal election, including the election outcomes, voter turnout information, post-election survey and opinion polls. In Chapter 3, we explore the data of CCHS conducted between the year 2007 to 2014. We identify the significant socio-demographics that are associated with life satisfaction status among CCHS respondents. The analysis results indicate that male respondents are less likely to feel satisfied with their lives compare to female respondents. Moreover, compare to the youngest group (age 18-24), the respondents in other age groups are less likely to feel satisfied. In terms of the highest education level, respondents with a post-secondary degree tend to feel satisfied. A positive relation shows between life satisfaction and annual household income, i.e., respondents with higher household income are more likely to feel satisfied.

In Chapter 4, we propose the joint models to link Canadian federal election outcomes with the longitudinal measurements of life satisfaction and other socio-demographics among Canadians. The joint models consist of two models that are linked to each other through the shared random slope. We apply two estimation approaches to obtain the parameter estimates, a two-stage estimation approach and a maximum likelihood estimation approach by applying the MCEM algorithm. The two approaches produce similar estimation results. The two-stage approach is easier to be carried out, whereas applying the MCEM algorithm is relatively time-consuming. In terms of the estimation accuracy, the MCEM algorithm is a maximum likelihood based approach which provides better estimates compare to the two-stage approach. The two-stage approach is merely the first iteration of the MCEM algorithm, and the estimates from the first iteration often deviate from the true value.

We use voter information only, CCHS respondents information only, or combine these two data sources to model the election outcomes, and obtain different estimation results. The estimation results indicate that the changing in life satisfaction plays an important role in the federal election, the ridings with a higher random slope for life satisfaction tend to have a higher probability of voting for a new political party that is different from the previous election. The results also show that male voters tend to vote for the same political party, and the young adults (age 18-34) are more likely to vote for a new political party. The effects of the highest education level and the annual household income should be interpreted with caution. A possible correlation between these two factors may exist, i.e., people with a higher level of education are more likely to have a higher income.

In Chapter 5, we present two examples of applying the proposed joint models. We define a different primary response variable as the indicator of the winning party in each riding changes from other political parties to NDP. The two-stage estimation results show that male voters are less likely to change to vote for NDP, whereas young adults are more likely to do so. Also, the ridings that have a higher random slope for life satisfaction tend to have a higher probability of changing to vote for NDP. In another application of the joint models, we predict the outcomes for the upcoming Canadian federal election in October 2019. After comparing different scenarios under each estimation methods, we predict that the Liberals are more likely to form a majority government compares to the Conservatives.

6.2 Future Investigations

In the longitudinal model of life satisfaction, we assume that the percentage of respondents feel satisfied is independent among the ridings across Canada. However, the life satisfaction status of one riding is highly likely to be correlated with the life satisfaction status in nearby ridings. We plan on investigating the possible spatial correlation among the ridings. We also assume for each riding, the random error for different years are independent and follow an identical normal distribution with mean zero. This may not be the case; that is, the random error may be dependent. Therefore, the possible temporal correlation will be considered for our further investigation.

Moreover, to further explore how life satisfaction level and other socio-demographics among Canadians affect the federal election outcomes, we consider conducting a simulation study in the future. For example, we could increase the percentage of young adults among the voters for some ridings to see how the increase will change the election outcomes.

To better predict the election outcomes, we plan on computing prediction intervals based on the variance estimation of the estimators for the parameters in the proposed joint modeling. In the election outcomes model, we propose four estimation methods to estimate the covariates of voters on riding level. These estimates may cause measurement errors. Further investigation and discussion will be carried out to evaluate the potential measurement errors and how to minimize the errors. To further compare these four estimation methods, different measures will be provided and discussed to evaluate the goodness of fit of the joint models.

Last but not least, in the proposed joint models, we only use the election outcomes and voter turnout information as the election data sources. Future investigations plan on including information from the opinion polls and the post-election survey. The results of multiple opinion polls conducted before the election day could be included to make a dynamic prediction of the election outcomes. Also, the target population of CCHS is not equivalent to the population of voters. Part of the CCHS target population is the electors who could vote in the election but choose not to. The post-election survey provides information on the proportion of sampled electors who did not vote and the corresponding reasons. We plan on making use of this part of information to adjust the difference between the CCHS target population and the electors in Canadian federal election.

Bibliography

Agresti, A. *Categorical Data Analysis*. Wiley, 2014.

Bélanger, É. & Godbout, J.-F. Forecasting canadian federal elections. *PS: Political Science and Politics*, 43(4):691–699, 2010. doi: 10.1017/S1049096510001113.

Campbell, J. E. Forecasting the presidential vote in the states. *American Journal of Political Science*, 36(2):386, 1992. doi: 10.2307/2111483.

Clarke, H. D., Gravelle, T. B., Scotto, T. J., Stewart, M. C. & Reifler, J. Like father, like son: Justin trudeau and valence voting in canada's 2015 federal election. *PS: Political Science and Politics*, 50(03):701–707, 2017. doi: 10.1017/s1049096517000452.

Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977. doi: 10.1111/j.2517-6161.1977.tb01600.x.

Elections Canada. Estimation of voter turnout by age group, Aug 2004. URL `https://www.elections.ca/content.aspx?section=res&dir=rec/part/estim&document=index&lang=e`.

Elections Canada. Survey of electors following the 41st general election, 2011. URL `https://www.elections.ca/content.aspx?section=res&dir=rec/eval/pes2011/elsvy&document=index&lang=e`. [Online; accessed 09-July-2019].

Elections Canada. `https://www.elections.ca/home.aspx`, 2019. [Online; accessed 09-July-2019].

Horrocks, J. & Heuvel, M. J. Prediction of pregnancy: a joint model for longitudinal and binary data. *Bayesian Analysis*, 4(3):523–538, 2009. doi: 10.1214/09-ba419.

Kleinbaum, D. G., Kupper, L. L., Nizam, A. & Rosenberg, E. S. *Applied regression analysis and other multivariable methods*. Cengage Learning, 2014.

Levine, R. A & Casella, G. Implementations of the monte carlo em algorithm. *Journal of Computational and Graphical Statistics*, 10(3):422–439, 2001. doi: 10.1198/106186001317115045. URL `https://doi.org/10.1198/106186001317115045`.

Lombardo, P., Jones, W., Wang, L., Shen, X. & Goldner, E. M. The fundamental association between mental health and life satisfaction: results from successive waves of a canadian national survey. *BMC Public Health*, 18(1), 2018. doi: 10.1186/s12889-018-5235-x.

McCulloch, C. E. Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, 92(437):162–170, 1997. doi: 10.1080/01621459.1997.10473613. URL https://doi.org/10.1080/01621459.1997.10473613.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2017. URL https://www.R-project.org/.

Rosenthal, J. S. Was the conservative majority predictable? *Canadian Journal of Statistics*, 39(4):721–733, 2011. doi: 10.1002/cjs.10126.

SAS Institute Inc. *SAS/STAT Software, Version 9.4.* Cary, NC, 2014. URL http://www.sas.com/.

SAS Institute Inc. 2008. *SAS/STAT® 9.2 Users Guide.* Cary, NC: SAS Institute Inc.

Statistics Canada. How's life in the city? life satisfaction across census metropolitan areas and economic regions in canada, Nov 2015. URL https://www150.statcan.gc.ca/n1/pub/11-626-x/11-626-x2015046-eng.htm.

Statistics Canada. Canadian community health survey - annual component (cchs), November 2018. URL http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=795204.

Terry Hei Wai Tang. Predictive estimation in canadian federal elections, 2017. URL https://www.stat.sfu.ca/content/dam/sfu/stat/alumnitheses/2017/Thesis-Terry%20Tang.pdf.

Ward, N. & Courtney, J. Canadian elections, 2018. URL https://www.thecanadianencyclopedia.ca/en/article/elections.

Wikipedia contributors. Opinion polling for the 2015 canadian federal election, Jun 2019. URL https://en.wikipedia.org/wiki/Opinion_polling_for_the_2015_Canadian_federal_election.

Appendix A

List of Tables

Table 1: The Distribution of Seats in Canadian Federal Election by Political Party.

Political Party	**2004 count (%)**	**2006 count (%)**	**2008 count (%)**	**2011 count (%)**	**2015 count (%)**
Conservatives	99 (32.14)	124 (40.26)	143 (46.43)	166 (53.90)	99 (29.29)
Liberals	135 (43.83)	103 (33.44)	77 (25.00)	34 (11.04)	184 (54.44)
NDP	19 (6.17)	29 (9.42)	37 (12.01)	103 (33.44)	44 (13.02)
Bloc Québécois	54 (17.53)	51 (16.56)	49 (15.91)	4 (1.30)	10 (2.96)
Others	1 (0.32)	1 (0.32)	2 (0.65)	1 (0.32)	1 (0.30)

Table 2: Summary of the Opinion Polls Results for the 42^{nd} Canadian Federal Election (in percentage).

Polling Company	**Polling Date**	**Con.**	**Lib.**	**NDP**	**BQ**	**Green**	**Margin of Error (+/-)**	**Sample Size**
Nanos	18 Oct 2015	30.5	39.1	19.7	5.5	4.6	3.5	800
Forum	18 Oct 2015	30	40	20	6	3	3	1373
Mainstreet	18 Oct 2015	32	39	21	3	5	1.4	5063
EKOS	18 Oct 2015	31.9	35.8	20.4	4.9	5.6	2.1	2122
EKOS	17 Oct 2015	32.6	34.4	21.0	5.4	5.4	2.4	1621
Ipsos Reid	17 Oct 2015	31	38	22	4	4	2.2	2503
Nanos	17 Oct 2015	30.5	37.3	22.1	4.6	4.7	2.2	2000

Table 3: The Distribution of Life Satisfaction Level among CCHS Respondents.

Life Satisfiaction (%)	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
Satisfied or Very Satisfied	89.09	88.84	87.71	88.08	88.13	87.77
Neither Satisfied Nor Dissatisfied	5.07	5.29	6.06	6.00	5.65	5.80
Dissatisfied or Very Dissatisfied	3.41	3.39	3.37	3.08	3.01	2.96
NA	2.43	2.48	2.86	2.84	3.20	3.47

Life Satisfiaction (%)	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
Satisfied or Very Satisfied	87.85	87.79	88.25	88.13	88.16	88.38
Neither Satisfied Nor Dissatisfied	5.63	5.55	5.31	5.19	5.35	5.11
Dissatisfied or Very Dissatisfied	3.09	3.11	3.17	3.39	3.28	3.26
NA	3.42	3.55	3.28	3.30	3.21	3.26

Table 4: Survey of Electors Following the 41^{st} General Election: Summary of Socio-demographic Information.

Social-demographic Characteristics	**Percentage (N = 3570)**
Gender	
Male	49%
Female	51%
Age	
18 - 24	12%
25 - 44	35%
45 - 64	34%
65+	17%
NA	2%
Education Level	
High school diploma or less	31%
Community college/vocational/trade school/commercial/CEGEP	29%
Some university	6%
University degree	33%
NA	1%
Household Annual Income (CAD)	
<20,000	8%
20,000 - 40,000	15%
40,000 - 60,000	15%
60,000 - 80,000	14%
80,000 - 100,000	11%
>= 100,000	23%
NA	13%

Table 5: The Logistic Regression Analysis with CCHS in Year 2011

Estimate Coefficients	**Estimates**	**Standard Error**	**P-val**
Intercept	1.9129	0.0800	<.0001
Gender (ref: Female)			
Male	-0.3070	0.0268	<.0001
Age (ref: 18-24)			
25 - 34	-0.1662	0.0721	0.0211
35 - 44	-0.5406	0.0692	<.0001
45 - 54	-0.7631	0.0651	<.0001
55 - 64	-0.6361	0.0631	<.0001
65 - 74	-0.5795	0.0647	<.0001
75 +	-1.0891	0.0632	<.0001
Education (ref: No Post-secondary Degree)			
Certificate or Diploma	0.2092	0.0628	0.0009
Bachelor or Above	0.4517	0.0718	<.0001
NA	-0.1167	0.0609	0.0554
Household Income (CAD) (ref: <20,000)			
20,000 - 40,000	0.4655	0.0372	<.0001
40,000 - 60,000	0.8114	0.0428	<.0001
60,000 - 80,000	1.0765	0.0503	<.0001
80,000 - 10,000	1.2898	0.0629	<.0001
>= 100,000	1.5306	0.0538	<.0001

Table 6: Two-stage Estimation: Life Satisfaction Model with Young Adults Group

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	1.8842	0.0476	<.0001
Year	0.1867	0.0582	0.0014
Male	-0.0026	0.0122	0.8300
Young Adults	0.0393	0.0144	0.0064
Certificate/Diploma	0.0359	0.0141	0.0111
Income <40,000	-0.0540	0.0152	0.0004

Covariance	**Estimate**	**Standard Error**	**P-val**
Random Intercept	0.0519	0.0120	<.0001
Random Slope	0.0219	0.0183	0.1156
Residual	0.0781	0.0047	<.0001

Table 7: Two-stage Estimation : Election Outcomes Model with Young Adults Group - Predictors Estimation Method 1

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	-1.2407	0.1874	<.0001
Changing in LS	3.6321	2.4804	0.1431
Male	-2.2208	0.4066	<.0001
Young Adults	2.0350	0.3527	<.0001

Table 8: Two-stage Estimation : Election Outcomes Model with Young Adults Group - Predictors Estimation Method 2

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	-0.7730	0.1284	<.0001
Changing in LS	9.2784	2.3231	<.0001
Male	-0.2813	0.1286	0.0287
Young Adults	0.2303	0.1276	0.0712

Table 9: Two-stage Estimation : Election Outcomes Model with Young Adults Group - Predictors Estimation Method 3

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	-0.7801	0.1294	<.0001
Changing in LS	9.1700	2.3383	<.0001
Male	-0.3104	0.1295	0.0165
Young Adults	0.2900	0.1290	0.0246

Table 10: Two-stage Estimation : Election Outcomes Model with Young Adults Group - Predictors Estimation Method 4

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	-0.7983	0.1317	<.0001
Changing in LS	7.5136	1.6692	<.0001
Male	-0.3192	0.1312	0.0150
Young Adults	0.2578	0.1315	0.0499

Table 11: MCEM Estimation : Joint Models with Young Adults Group - Predictors Estimation Method 1

Life Satisfaction Model

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	1.8846	0.0403	<.0001
Year	0.1865	0.0512	0.0003
Male	-0.0016	0.0102	0.8754
Young Adults	0.0377	0.0099	0.0001
Certificate/Diploma	0.0359	0.0096	0.0002
Income <40,000	-0.0540	0.0105	<.0001

Covariance	**Estimate**	**Standard Error**	**P-val**
Random Intercept	0.0506	0.0041	<.0001
Random Slope	0.0231	0.0019	<.0001
Residual	0.0777	0.0036	<.0001

Election Outcomes Model

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	-1.2808	0.1881	<.0001
Changing in LS	0.7217	0.9095	0.4275
Male	-2.3745	0.3983	<.0001
Young Adults	2.1843	0.3422	<.0001

Table 12: MCEM Estimation : Joint Models with Young Adults Group - Predictors Estimation Method 2

Life Satisfaction Model

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	1.8836	0.0403	<.0001
Year	0.1858	0.0512	0.0003
Male	-0.0021	0.0102	0.8369
Young Adults	0.0403	0.0099	<.0001
Certificate/Diploma	0.0353	0.0096	0.0002
Income <40,000	-0.0543	0.0105	<.0001

Covariance	**Estimate**	**Standard Error**	**P-val**
Random Intercept	0.0508	0.0041	<.0001
Random Slope	0.0228	0.0018	<.0001
Residual	0.0776	0.0036	<.0001

Election Outcomes Model

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	-0.7473	0.1248	<.0001
Changing in LS	1.4276	0.8253	0.0837
Male	-0.2592	0.1237	0.0361
Young Adults	0.2455	0.1227	0.0454

Table 13: MCEM Estimation : Joint Models with Young Adults Group - Predictors Estimation Method 3

Life Satisfaction Model

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	1.8836	0.0403	<.0001
Year	0.1858	0.0512	0.0003
Male	-0.0021	0.0102	0.8369
Young Adults	0.0403	0.0099	<.0001
Certificate/Diploma	0.0353	0.0096	0.0002
Income <40,000	-0.0543	0.0105	<.0001

Covariance	**Estimate**	**Standard Error**	**P-val**
Random Intercept	0.0508	0.0041	<.0001
Random Slope	0.0228	0.0018	<.0001
Residual	0.0776	0.0036	<.0001

Election Outcomes Model

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	-0.7559	0.1259	<.0001
Changing in LS	1.4003	0.8309	0.0919
Male	-0.2845	0.1245	0.0223
Young Adults	0.3154	0.1244	0.0112

Table 14: MCEM Estimation : Joint Models with Young Adults Group - Predictors Estimation Method 4

Life Satisfaction Model

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	1.8836	0.0403	<.0001
Year	0.1858	0.0512	0.0003
Male	-0.0021	0.0102	0.8369
Young Adults	0.0403	0.0099	<.0001
Certificate/Diploma	0.0353	0.0096	0.0002
Income <40,000	-0.0543	0.0105	<.0001

Covariance	**Estimate**	**Standard Error**	**P-val**
Random Intercept	0.0508	0.0041	<.0001
Random Slope	0.0228	0.0018	<.0001
Residual	0.0776	0.0036	<.0001

Election Outcomes Model

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	-0.7588	0.1262	<.0001
Changing in LS	1.3872	0.8324	0.0956
Male	-0.3108	0.1258	0.0135
Young Adults	0.3090	0.1250	0.0134

Table 15: Two-stage Estimation: Life Satisfaction Model with Ratio of Young over Middle-aged Adults

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	1.8375	0.0586	<.0001
Year	0.1958	0.0584	0.0008
Male	-0.0015	0.0123	0.9038
Young/Middle-Aged Adults	0.0946	0.0800	0.2372
Certificate/Diploma	0.0340	0.0142	0.0170
Income <40,000	-0.0586	0.0152	0.0001

Covariance	**Estimate**	**Standard Error**	**P-val**
Random Intercept	0.0526	0.0121	<.0001
Random Slope	0.0229	0.0185	0.1085
Residual	0.0783	0.0047	<.0001

Table 16: Two-stage Estimation : Election Outcomes Model with Ratio of Young over Middle-aged Adults - Predictors Estimation Method 1

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	-19.0850	3.2121	<.0001
Changing in LS	4.2713	2.3893	0.0738
Male	-1.8059	0.3449	<.0001
Young/Middle-Aged Adults	51.1737	8.8594	<.0001

Table 17: Two-stage Estimation : Election Outcomes Model with Ratio of Young over Middle-aged Adults - Predictors Estimation Method 2

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	-1.0311	0.4309	0.0167
Changing in LS	9.4225	2.2521	<.0001
Male	-0.2724	0.1290	0.0347
Young/Middle-Aged Adults	0.5259	0.8257	0.5242

Table 18: Two-stage Estimation : Election Outcomes Model with Ratio of Young over Middle-aged Adults - Predictors Estimation Method 3

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	-1.4203	0.4410	0.0013
Changing in LS	9.2397	2.2704	<.0001
Male	-0.3021	0.1301	0.0202
Young/Middle-Aged Adults	1.7355	1.1226	0.1221

Table 19: Two-stage Estimation : Election Outcomes Model with Ratio of Young over Middle-aged Adults - Predictors Estimation Method 4

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	-1.4281	0.4477	0.0014
Changing in LS	7.6491	1.6651	<.0001
Male	-0.3155	0.1317	0.0166
Young/Middle-Aged Adults	1.7081	1.1402	0.1341

Table 20: MCEM Estimation : Joint Models with Ratio of Young over Middle-aged Adults - Predictors Estimation Method 1

Life Satisfaction Model

Parameter	Estimate	Standard Error	P-val
Intercept	1.8336	0.0471	<.0001
Year	0.1968	0.0512	0.0001
Male	-0.0019	0.0102	0.8522
Young/Middle-Aged Adults	0.0988	0.0577	0.0868
Certificate/Diploma	0.0344	0.0097	0.0004
Income <40,000	-0.0596	0.0103	<.0001

Covariance	Estimate	Standard Error	P-val
Random Intercept	0.0519	0.0042	<.0001
Random Slope	0.0230	0.0018	<.0001
Residual	0.0779	0.0036	<.0001

Election Outcomes Model

Parameter	Estimate	Standard Error	P-val
Intercept	-20.7384	3.1640	<.0001
Changing in LS	0.6428	0.9036	0.4769
Male	-1.9559	0.3443	<.0001
Young/Middle-Aged Adults	55.7767	8.7071	<.0001

Table 21: MCEM Estimation : Joint Models with Ratio of Young over Middle-aged Adults - Predictors Estimation Method 2

Life Satisfaction Model

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	1.8375	0.0470	<.0001
Year	0.1979	0.0511	0.0001
Male	-0.0009	0.0101	0.9290
Young/Middle-Aged Adults	0.0927	0.0576	0.1075
Certificate/Diploma	0.0341	0.0097	0.0004
Income <40,000	-0.0597	0.0103	<.0001

Covariance	**Estimate**	**Standard Error**	**P-val**
Random Intercept	0.0521	0.0042	<.0001
Random Slope	0.0233	0.0019	<.0001
Residual	0.0777	0.0036	<.0001

Election Outcomes Model

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	-1.0783	0.4143	0.0092
Changing in LS	1.4025	0.8122	0.0842
Male	-0.2462	0.1237	0.0466
Young/Middle-Aged Adults	0.6858	0.7896	0.3851

Table 22: MCEM Estimation : Joint Models with Ratio of Young over Middle-aged Adults - Predictors Estimation Method 3

Life Satisfaction Model

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	1.8375	0.0470	<.0001
Year	0.1979	0.0511	0.0001
Male	-0.0009	0.0101	0.9290
Young/Middle-Aged Adults	0.0927	0.0576	0.1075
Certificate/Diploma	0.0341	0.0097	0.0004
Income <40,000	-0.0597	0.0103	<.0001

Covariance	**Estimate**	**Standard Error**	**P-val**
Random Intercept	0.0521	0.0042	<.0001
Random Slope	0.0233	0.0019	<.0001
Residual	0.0777	0.0036	<.0001

Election Outcomes Model

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	-1.5236	0.4254	0.0003
Changing in LS	1.3638	0.8178	0.0954
Male	-0.2721	0.1247	0.0291
Young/Middle-Aged Adults	2.0941	1.0770	0.0518

Table 23: MCEM Estimation : Joint Models with Ratio of Young over Middle-aged Adults - Predictors Estimation Method 4

Life Satisfaction Model

Parameter	Estimate	Standard Error	P-val
Intercept	1.8363	0.0471	<.0001
Year	0.1989	0.0512	0.0001
Male	-0.0012	0.0102	0.9063
Young/Middle-Aged Adults	0.0931	0.0577	0.1066
Certificate/Diploma	0.0337	0.0097	0.0005
Income <40,000	-0.0596	0.0103	<.0001

Covariance	Estimate	Standard Error	P-val
Random Intercept	0.0525	0.0042	<.0001
Random Slope	0.0228	0.0018	<.0001
Residual	0.0780	0.0036	<.0001

Election Outcomes Model

Parameter	Estimate	Standard Error	P-val
Intercept	-1.5003	0.4274	0.0004
Changing in LS	1.5170	0.8318	0.0682
Male	-0.3085	0.1262	0.0145
Young/Middle-Aged Adults	2.0214	1.0824	0.0618

Table 24: Change to NDP: Election Outcomes Model with Young Adults Group Two-stage Estimation - Predictors Estimation Method 1

Parameter	Estimate	Standard Error	P-val
Intercept	-2.6039	0.3227	<.0001
Changing in LS	8.1578	3.3379	0.0145
Male	-3.5290	0.6246	<.0001
Young Adults	3.4201	0.5126	<.0001

Table 25: Change to NDP: Election Outcomes Model with Young Adults Group Two-stage Estimation - Predictors Estimation Method 2

Parameter	Estimate	Standard Error	P-val
Intercept	-1.4620	0.1613	<.0001
Changing in LS	16.8702	2.9750	<.0001
Male	-0.1951	0.1491	0.1909
Young Adults	0.2656	0.1514	0.0795

Table 26: Change to NDP: Election Outcomes Model with Young Adults Group Two-stage Estimation - Predictors Estimation Method 3

Parameter	Estimate	Standard Error	P-val
Intercept	-1.4826	0.1637	<.0001
Changing in LS	16.7515	2.9912	<.0001
Male	-0.2383	0.1498	0.1117
Young Adults	0.3731	0.1533	0.0150

Table 27: Change to NDP: Election Outcomes Model with Young Adults Group Two-stage Estimation - Predictors Estimation Method 4

Parameter	Estimate	Standard Error	P-val
Intercept	-1.4857	0.1641	<.0001
Changing in LS	16.6370	2.9873	<.0001
Male	-0.2243	0.1495	0.1334
Young Adults	0.3816	0.1537	0.0130

Table 28: Change to NDP: Election Outcomes Model with Ratio of Young over Middle-aged Adults Two-stage Estimation - Predictors Estimation Method 1

Parameter	Estimate	Standard Error	P-val
Intercept	-34.1712	4.9615	<.0001
Changing in LS	8.6472	3.2027	0.0069
Male	-2.9190	0.5519	<.0001
Young/Middle-Aged Adults	90.3064	13.4305	<.0001

Table 29: Change to NDP: Election Outcomes Model with Ratio of Young over Middle-aged Adults Two-stage Estimation - Predictors Estimation Method 2

Parameter	Estimate	Standard Error	P-val
Intercept	-2.5966	0.5386	<.0001
Changing in LS	16.6233	2.9189	<.0001
Male	-0.2176	0.1510	0.1494
Young/Middle-Aged Adults	2.9906	1.3207	0.0235

Table 30: Change to NDP: Election Outcomes Model with Ratio of Young over Middle-aged Adults Two-stage Estimation - Predictors Estimation Method 3

Parameter	Estimate	Standard Error	P-val
Intercept	-1.4203	0.4410	0.0013
Changing in LS	9.2397	2.2704	<.0001
Male	-0.3021	0.1301	0.0202
Young/Middle-Aged Adults	1.7355	1.1226	0.1221

Table 31: Change to NDP: Election Outcomes Model with Ratio of Young over Middle-aged Adults Two-stage Estimation - Predictors Estimation Method 4

Parameter	**Estimate**	**Standard Error**	**P-val**
Intercept	-2.6022	0.5404	<.0001
Changing in LS	16.4959	2.9105	<.0001
Male	-0.2096	0.1507	0.1643
Young/Middle-Aged Adults	3.0003	1.3251	0.0236

Table 32: The Prediction of 2019 Federal Election: The Distribution of Seats Change from the Previous Election in 2015 by Estimation Method.

Seats Distribution in 2015 Election		**Only Election**	**Only CCHS**	**Combine**	**Jointly Combine**
Liberals	181	62	56	54	48
Conservatives	99	21	30	29	28
NDP	44	16	19	19	19
Bloc Québécois	10	10	8	8	8
Green	1	0	1	1	1
Total	335	109	114	111	104

Appendix B

List of Figures

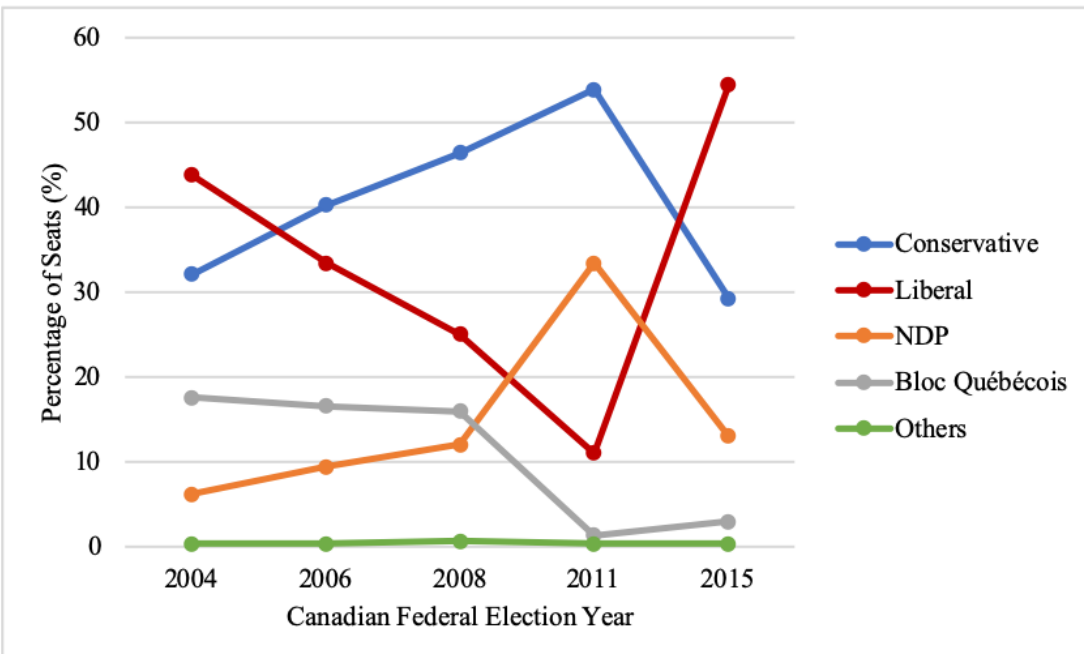

Figure 1: The Distribution of Seats in Canadian Federal Election by Political Party.

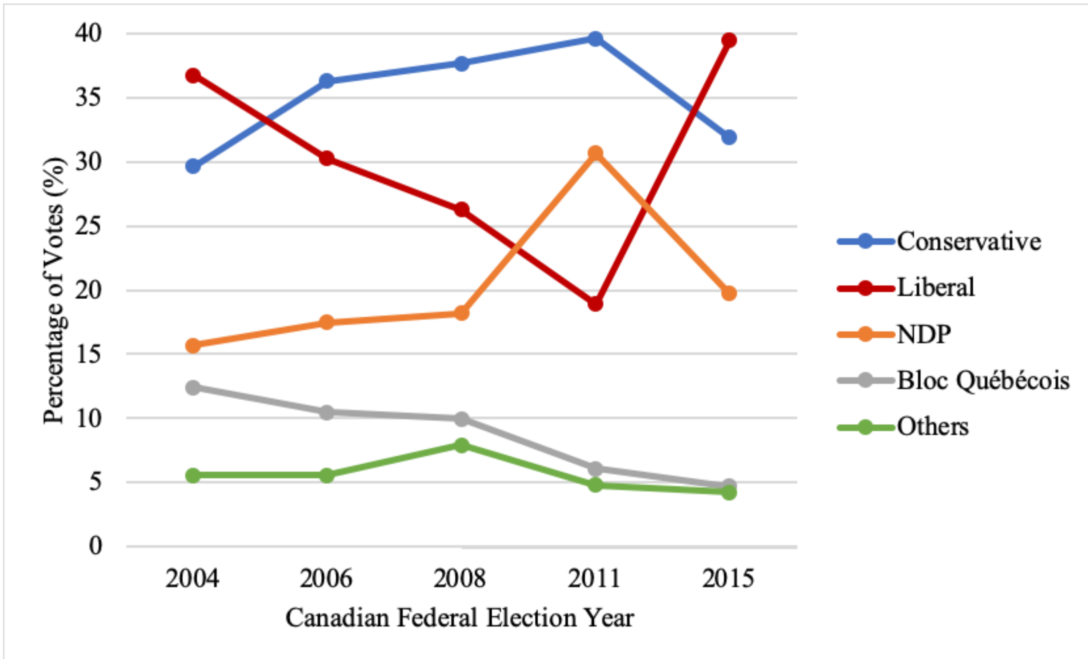

Figure 2: The Distribution of Votes in Canadian Federal Election by Political Party.

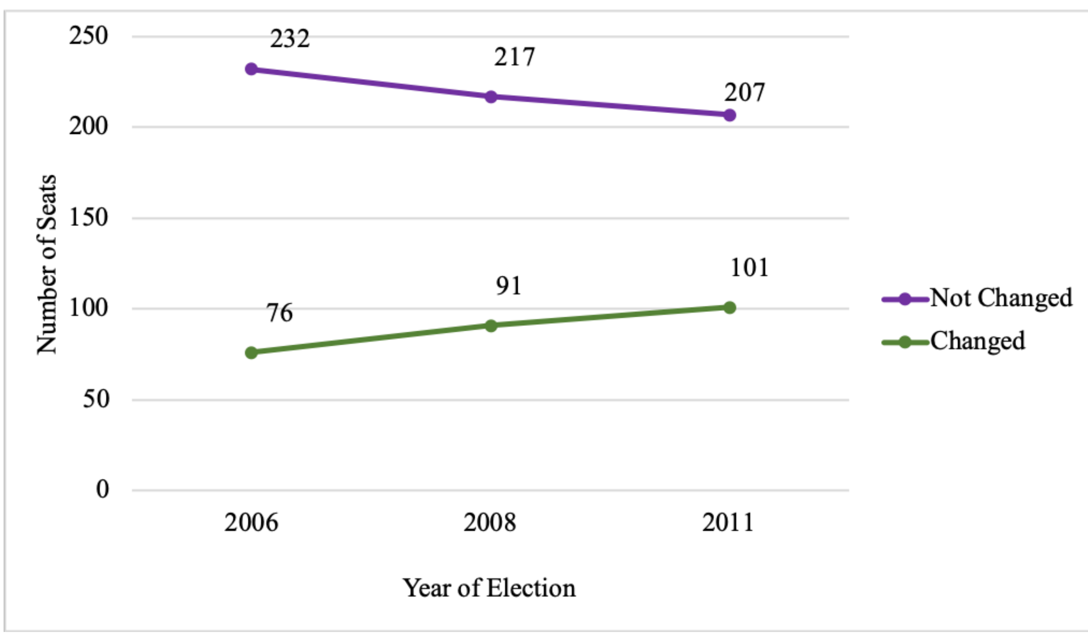

Figure 3: The Number of Seats in Canadian Federal Election: the Winning Political Party Changes from Previous Election

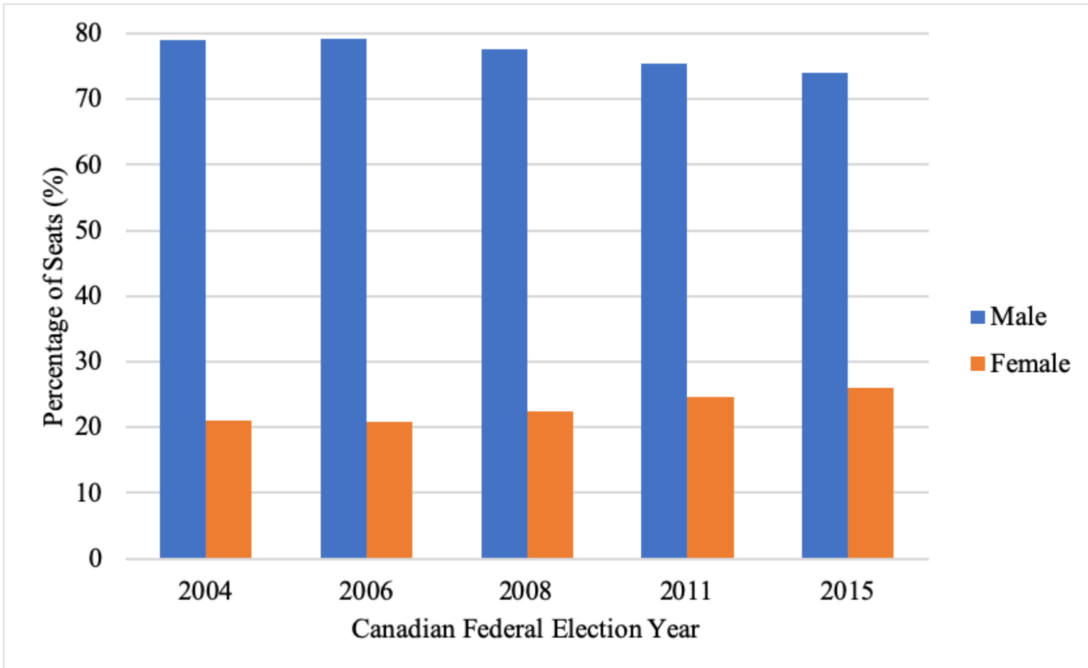

Figure 4: The Distribution of Seats in Canadian Federal Election by Elected MP's Gender.

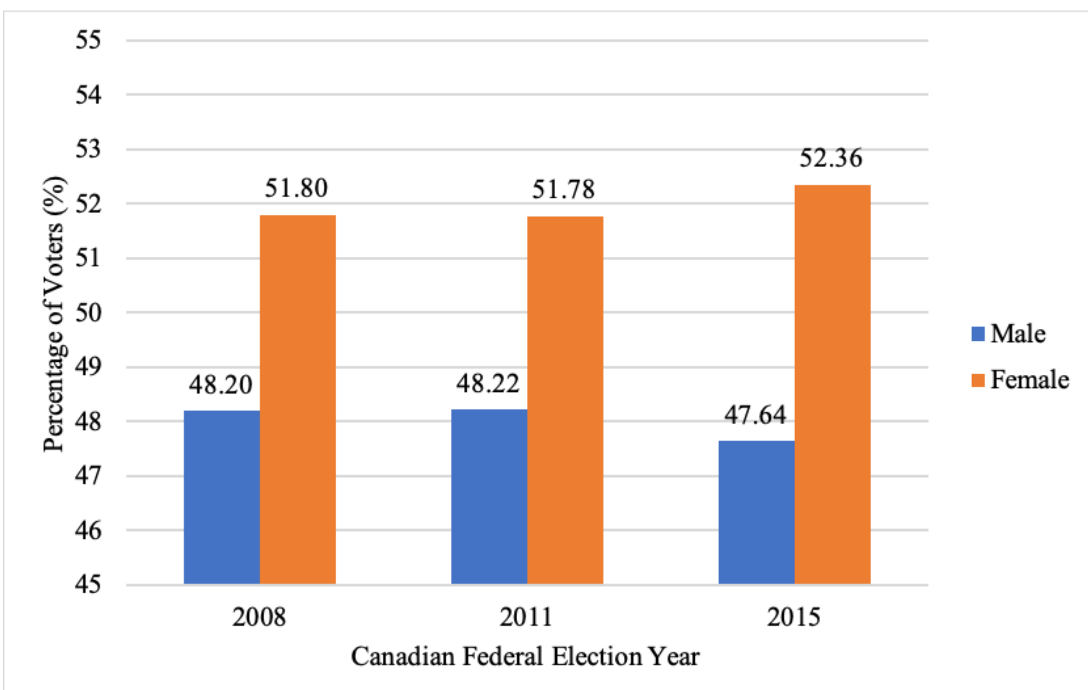

Figure 5: The Distribution of Voters in Canadian Federal Election by Gender.

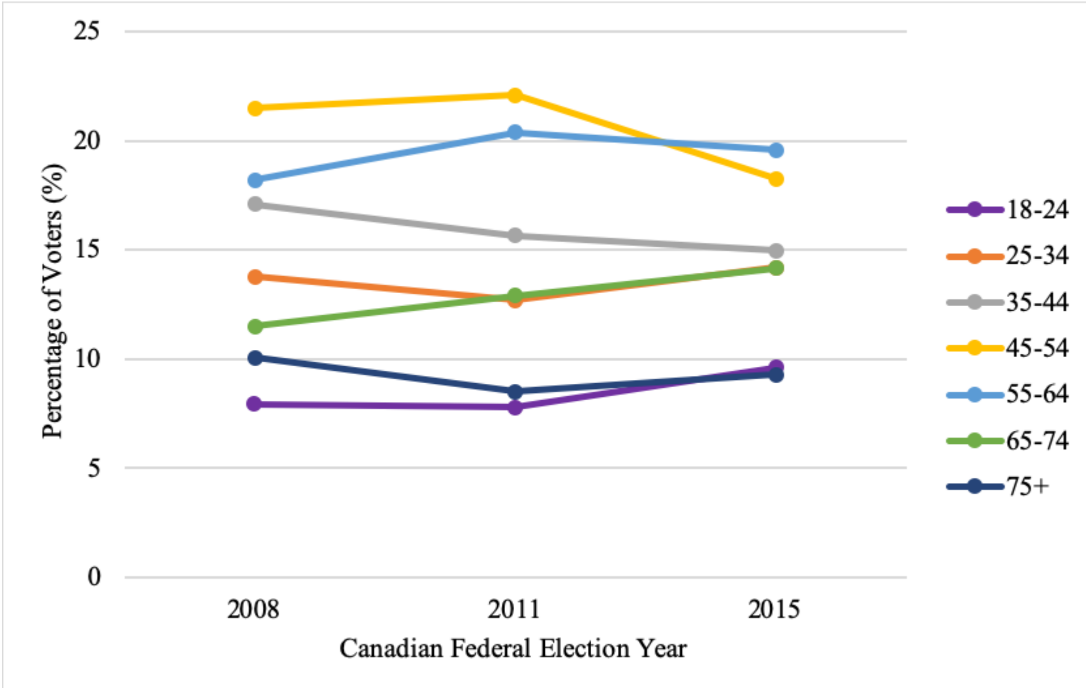

Figure 6: The Distribution of Voters in Canadian Federal Election by Age Groups.

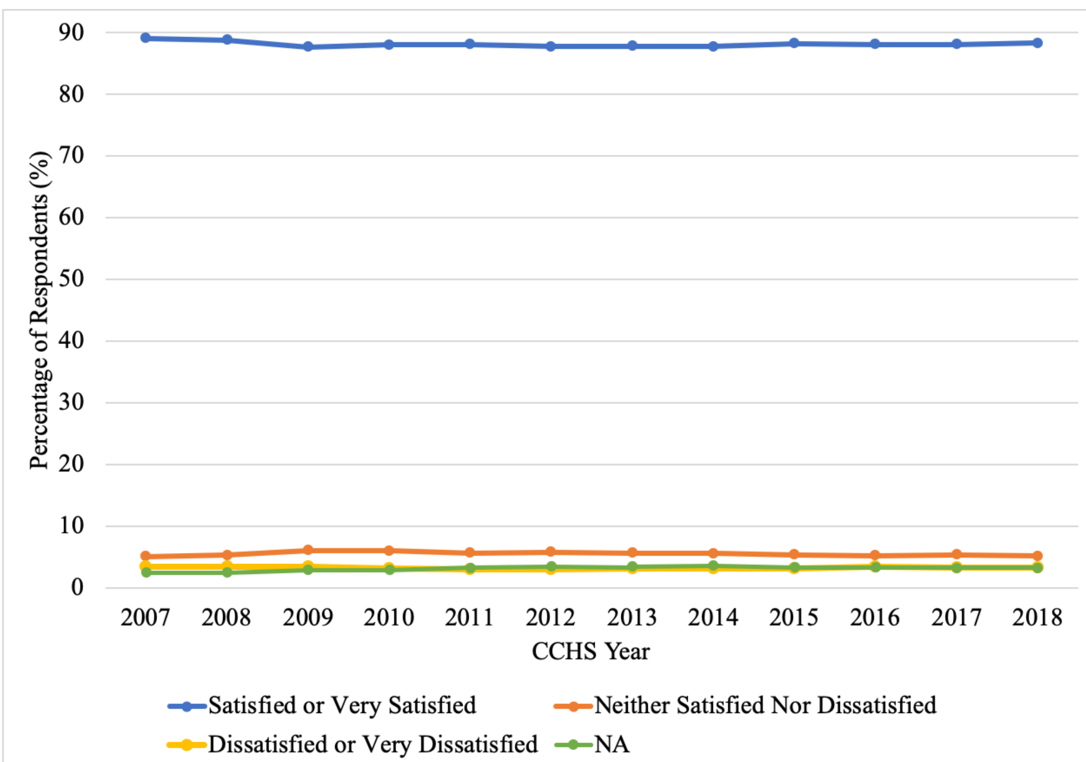

Figure 7: The Distribution of CCHS Respondents by Life Satisfaction Level.

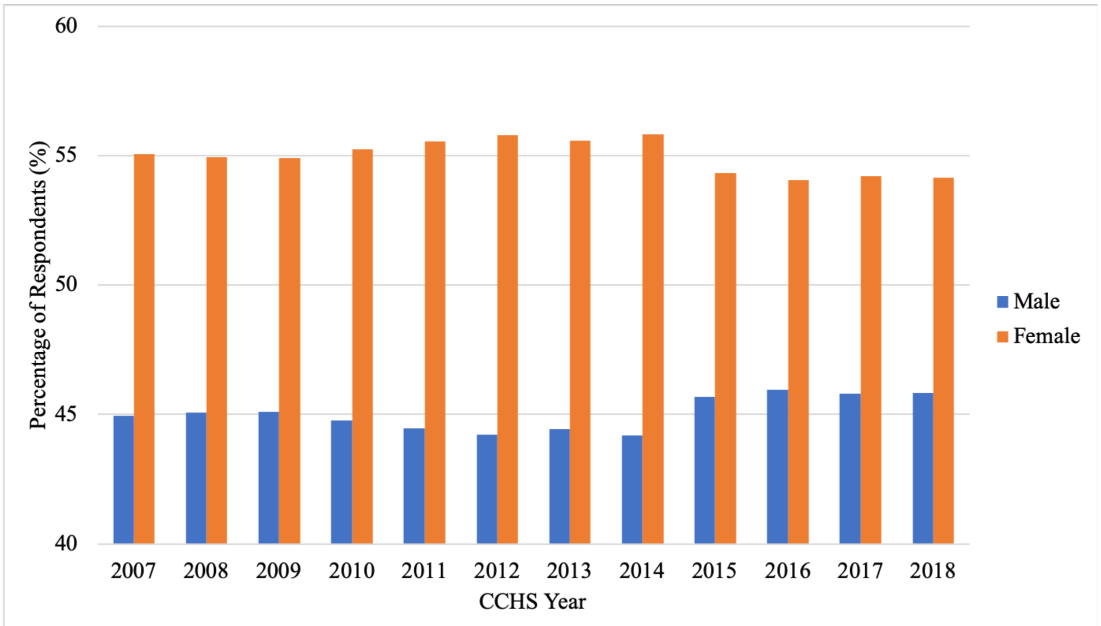

Figure 8: The Distribution of CCHS Respondents by Gender.

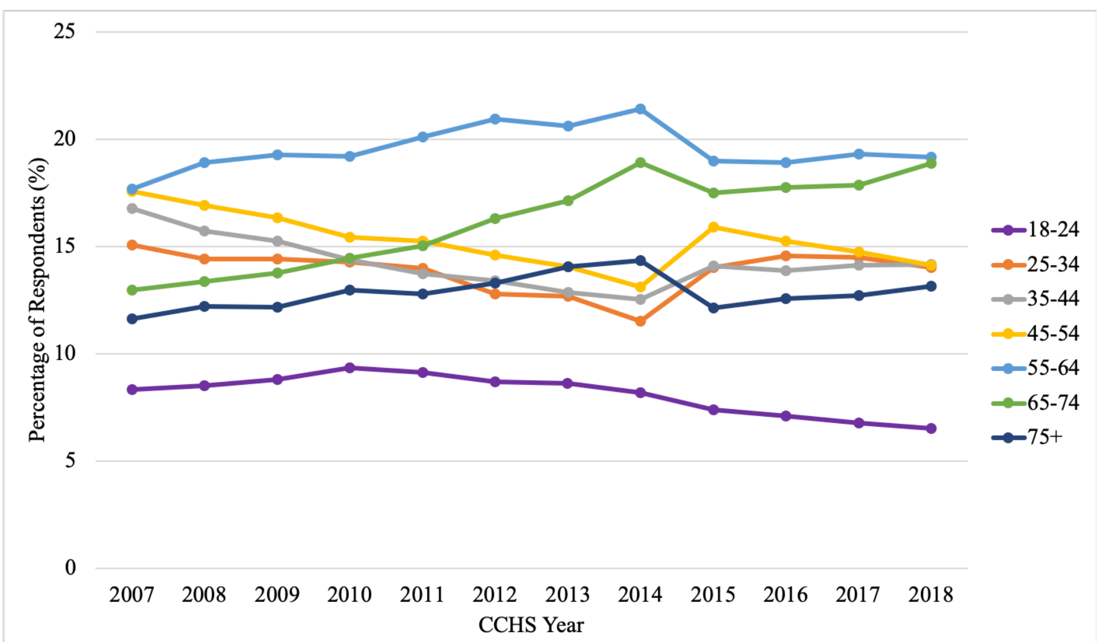

Figure 9: The Distribution of CCHS Respondents by Age Groups.

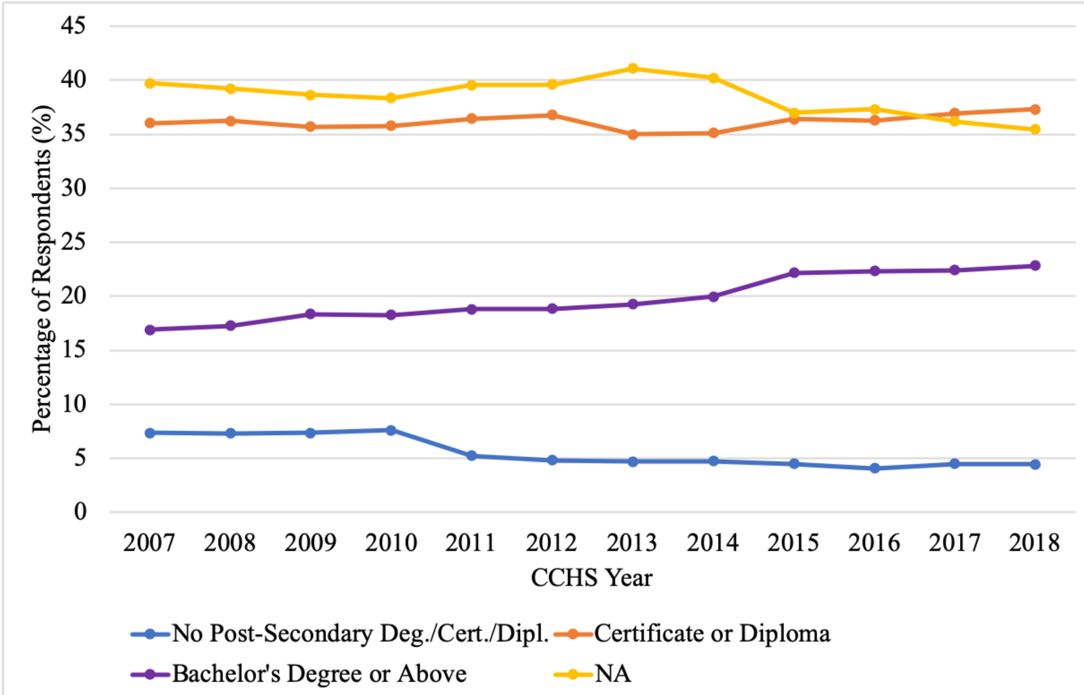

Figure 10: The Distribution of CCHS Respondents by Highest Education Level.

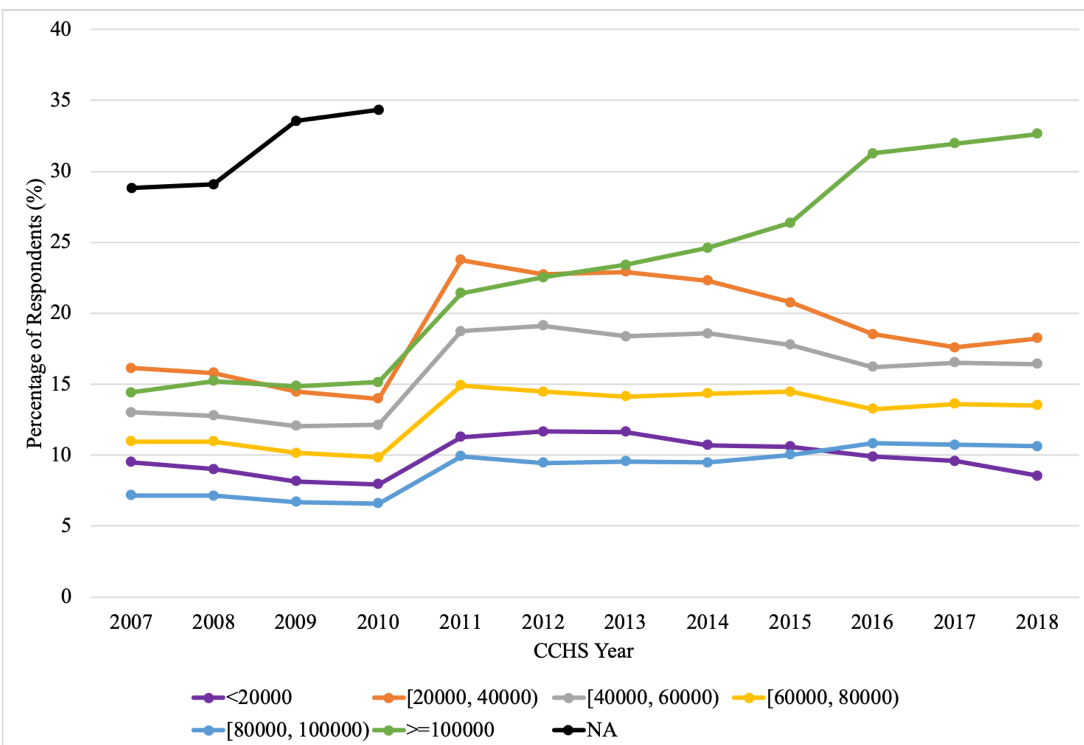

Figure 11: The Distribution of CCHS Respondents by Annual Household Income.

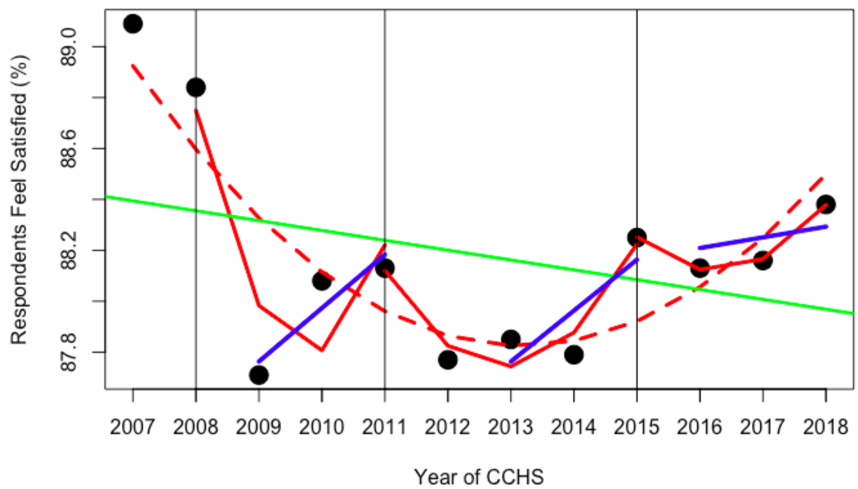

Figure 12: The Trend of CCHS Respondents Feel Satisfied or Very Satisfied with Lives between Year 2007 and 2018.

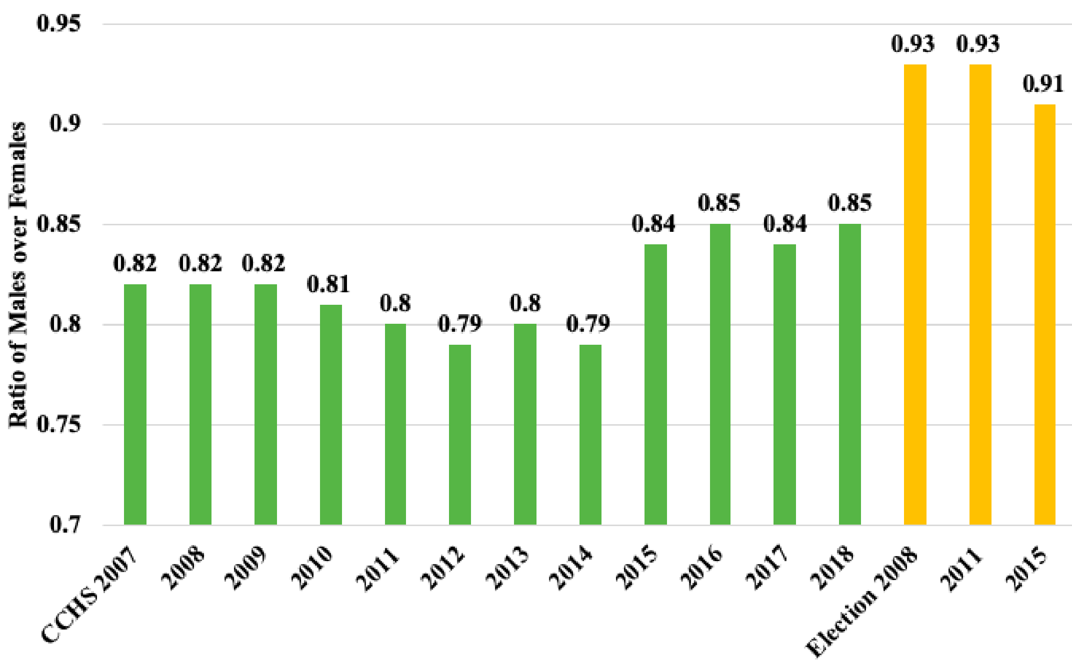

Figure 13: Comparing the Gender Ratio (Male over Female) between CCHS Respondents and Canadian Federal Election Voters.

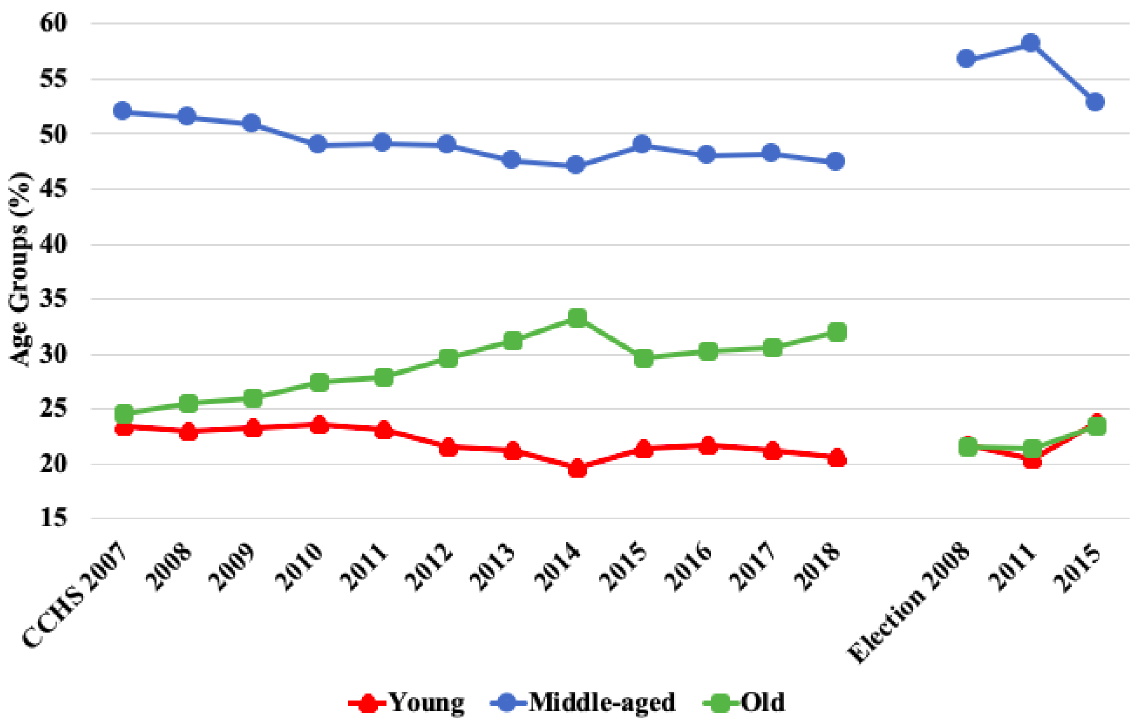

Figure 14: Comparing the Distribution of Age Groups between CCHS Respondents and Canadian Federal Election Voters.