RELATIONSHIPS BETWEEN INCOME INEQUALITY AND HEALTH: AN ECOLOGICAL CANADIAN STUDY.

by

Afshin Vafaei

A thesis submitted to the Department of Community Health & Epidemiology
in conformity with the requirements for
the degree of Master of Science

Queen’s University
Kingston, Ontario, Canada
(September, 2008)

Copyright ©Afshin Vafaei, 2008
Abstract

Background:

Many studies have demonstrated that health is a function of relative and not absolute income within populations. Canadian studies are not conclusive. There is a need for further investigation of the ‘relative income’ hypothesis in the Canadian population.

Objectives:

The primary objective of this research was to test the “relative income” hypothesis across Canadian health regions. The second objective was to extend the first hypothesis to consider rural versus urban populations in Canada.

Methods:

This research involved ecological analyses. The source of the data was the Canadian Community Health Survey, CCHS 2005 cycle 3.1. The units of analysis were health regions of Canada. Health of a region was estimated as the percentage of people who rated their health as good or excellent. The primary exposure variable was the ratio of people whose personal income was less than $15,000 relative to those reporting more than $80,000. Correlation analyses and multiple linear regressions were performed to ascertain the relationship between income inequality and health status in populations, adjusting for important covariates.

Results:

The measure of relative income inequality alone appeared to explain 18 per cent of the variability in the measure of health status in populations. However, after adding the measure of absolute income to the model, although 29 per cent of the variability was explained, the independent contribution of the inequality measure became non-significant. Linear regression models suggested that the absolute income variable alone could explain 30 per cent of the
variance in the health status of populations. Other variables with a statistically significant contribution to the final multiple regression model were education and alcohol consumption. Rural/urban status did not change the individual relationship between relative income inequality or absolute income and the measure of health status in populations.

**Conclusion:**

Across Canadian health regions, health status in populations was a function of absolute income but not relative income. Regions with higher levels of education had better levels of self-rated health. A larger percentage of heavy drinkers was also correlated with lower population health status. The study findings have implications for public health, economic policies, and social policies.
Co-Authorship

This thesis presents the research conducted by Afshin Vafaei in collaboration with his supervisors, Dr. Mark Rosenberg and Dr. William Pickett.

The idea of testing the ‘relative income’ hypothesis in Canada was Dr. Rosenberg’s. Afshin Vafaei developed procedures for extracting required data from the CCHS, 2005 in collaboration and under the supervision of Dr. Mark Rosenberg and Dr. William Pickett.

The development of the ecological models and methodological approaches, statistical analyses, interpretations of results and the writing of the manuscript were done by Afshin Vafaei with guidance, supervision, and editorial feedback provided by Dr. Rosenberg and Dr. Pickett.

The other components of the thesis (i.e., Introduction, Literature Review, General Discussion, and Appendices) were the work of Afshin Vafaei, who received suggestions with conceptual and editorial feedback from Drs. Rosenberg and Pickett.
Acknowledgements

I am indebted to my supervisors, Dr. Mark Rosenberg and Dr. William Pickett, for their continuous support and guidance. Their help was beyond a simple supervisory role. They were always available to answer my endless questions, provided timely constructive feedback, and they helped me not only in this project but in other aspects of academic life. Thank you!

I would like to thank all my classmates for being the most congenial group of people one could ever hope for. I learned a lot from them. I also appreciate the dedication of faculty and staff of department of Community Health and Epidemiology, Emergency Medicine research group, and other affiliated research groups I had the privilege of being a student at in different courses and programs.

My thanks to Lida for her love. She was most encouraging during my many months of work and was most helpful in distracting me from it.

And, finally, I thank Golab who graciously agreed not to sleep on the keyboard so I could type this thesis.
## Table of Contents

Abstract ................................................................................................................................. ii  
Co-Authorship ...................................................................................................................... iv  
Acknowledgements .............................................................................................................. v  
Table of Contents .................................................................................................................. vi  
List of Figures......................................................................................................................... x  
List of Tables ......................................................................................................................... xi  

Chapter 1 Introduction ........................................................................................................... 1  
  1.1 Background .................................................................................................................. 1  
  1.2 Objectives .................................................................................................................. 2  
  1.3 Overview of Study Design .......................................................................................... 3  
  1.4 Rationale ................................................................................................................... 4  
  1.5 Thesis Organization .................................................................................................. 5  
  1.6 References ................................................................................................................. 6  

Chapter 2 Literature Review ................................................................................................. 8  
  2.1 Literature Search Strategy ........................................................................................ 8  
  2.2 Social Determinants of Health ................................................................................. 8  
  2.3 Income Inequality and Health ................................................................................... 9  
    2.3.1 The Absolute Income Hypothesis ...................................................................... 9  
    2.3.2 The Relative Income Hypothesis ...................................................................... 10  
    2.3.3 Measures of Income Inequality ........................................................................ 12  
    2.3.4 Measures of Health ......................................................................................... 13  
  2.4 Potential Causal Mechanisms .................................................................................. 14  
    2.4.1 Individual Income Mechanisms ....................................................................... 15  
    2.4.2 Neo-material Mechanisms ............................................................................... 16  
    2.4.3 Social Capital Mechanisms ............................................................................. 16  
    2.4.4 Psycho-social Mechanisms ............................................................................. 17  
    2.4.5 Socio-biological Mechanisms ......................................................................... 17  
    2.4.6 Alternative Explanations .................................................................................. 18  
  2.5 Review of Studies on the Relationship between Income Inequality and Health ....... 18  
    2.5.1 International Studies ......................................................................................... 18  

vi
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5.1.1</td>
<td>Developed Countries</td>
<td>19</td>
</tr>
<tr>
<td>2.5.1.2</td>
<td>Developing Countries</td>
<td>20</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Within Countries Studies</td>
<td>21</td>
</tr>
<tr>
<td>2.5.2.1</td>
<td>Unites States Studies</td>
<td>21</td>
</tr>
<tr>
<td>2.5.2.2</td>
<td>Other Developed Countries</td>
<td>22</td>
</tr>
<tr>
<td>2.5.2.3</td>
<td>Canadian Studies</td>
<td>23</td>
</tr>
<tr>
<td>2.5.2.4</td>
<td>Developing Countries</td>
<td>25</td>
</tr>
<tr>
<td>2.6</td>
<td>Effect of Size of Regions on the Relationship of Interest</td>
<td>26</td>
</tr>
<tr>
<td>2.7</td>
<td>Effect of Rural/Urban Status on the Relationship of Interest</td>
<td>27</td>
</tr>
<tr>
<td>2.8</td>
<td>Methodological Issues</td>
<td>27</td>
</tr>
<tr>
<td>2.8.1</td>
<td>Artifactual Relationship between Income Distribution and Health</td>
<td>28</td>
</tr>
<tr>
<td>2.8.2</td>
<td>Ecological Fallacy/Atomistic Fallacy</td>
<td>28</td>
</tr>
<tr>
<td>2.8.3</td>
<td>Multi-level Approach</td>
<td>29</td>
</tr>
<tr>
<td>2.8.4</td>
<td>Choice of the Income Inequality Measure</td>
<td>29</td>
</tr>
<tr>
<td>2.9</td>
<td>Summary</td>
<td>30</td>
</tr>
<tr>
<td>2.10</td>
<td>References</td>
<td>31</td>
</tr>
</tbody>
</table>

Chapter 3 Methods
3.1 Overview
3.2 Purpose and Empirical Objectives
3.3 Ecological Model
3.4 Source of Data
3.4.1 Features of the CCHS cycle 3.1
3.4.1.1 Design
3.4.1.2 Weighting
3.4.1.3 Health regions
3.4.2 Access to CCHS cycle 3.1 data
3.5 Ascertainment of Outcome
3.6 Ascertainment of Exposure
3.7 Covariates
3.7.1 Absolute Income Variable
3.7.2 Health Care Utilization
3.7.3 Socio-demographic Variables
List of Figures

Figure 2.1 Theoretical Causal Mechanisms ........................................................................ 15
Figure 3.1 General Health Status Variable in the CCHS cycle 3.1 ............................... 49
Figure 3.2 Smoking Variable in the CCHS cycle 3.1 .................................................. 54
Figure 3.3 Social Capital Variable in the CCHS cycle 3.1 .......................................... 55
Figure B.1 Scatter plot of Population Health against Income Inequality .................... 98
Figure B.2 Scatter plot of Population Health against Absolute Income ...................... 99
Figure B.3 The Final Model ......................................................................................... 103
Figure B.4 Final Models in Rural and Urban Regions ................................................ 104
Figure B.5 Frequency distribution of measure of population health .......................... 107
Figure B.6 Frequency distribution of measure of Absolute Income ............................ 107
Figure B.7 Frequency distribution of measure of income inequality ......................... 108
Figure B.8 Logarithmic transformation of Income Inequality Variable ....................... 108
Figure B.9 Scatter plot of Population Health Status against Income Inequality in Male Regions ........................................................................................................... 109
Figure B.10 Scatter plot of Population Health Status against Income Inequality in Female Regions ........................................................................................................... 109
Figure B.11 Scatter plot of Population Health Status against Absolute Income in Male Regions ........................................................................................................... 110
Figure B.12 Scatter plot of Population Health Status against Absolute Income in Female Regions ........................................................................................................... 110
Figure B.13 Scatter plot of Population Health Status against Income Inequality in Senior Regions ........................................................................................................... 111
Figure B.14 Scatter plot of Population Health against Income Inequality in Non-senior Regions ........................................................................................................... 111
List of Tables

Table 3.1 Description of Dependent and Independent Variables ..................................................... 46
Table 3.2 Subgroups of Variables ........................................................................................................ 51

Manuscript
Table 1 Description of Variables ........................................................................................................ 79
Table 2 Mean, Minimum, Median, and Maximum Values of Variables Used in the analysis………….. 80
Table 3 Pearson Correlation Matrix of variables................................................................................. 81
Table 4 Results of linear regression analyses in different models ..................................................... 82
Table 5 Results of linear regression analysis of association between income inequality, absolute income and health within different levels of rural/urban status ............................. 83

Appendix B
Table B.1 Mean, Minimum, Median, and Maximum Values of Variables Used in the analysis………… 112
Table B.2 Pearson Correlation Matrix of Variables............................................................................. 113
Table B.3 Results of linear regression analysis of association between income inequality and health within different levels of covariates (Subgroup Analyses)......................................................... 114
Table B.4 Pearson coefficient and corresponding p-Values for the correlation between income inequality and health after controlling individually for other covariates........................................... 115
Table B.5 Results of linear regression analyses of association between income inequality and health, controlling individually for other covariates........................................................................... 116
Table B.6 Results of linear regression analysis of association between Absolute income and health within different levels of covariates (Subgroup Analyses)................................................................. 117
Table B.7 Pearson coefficient and corresponding p-Values for the correlation between absolute income and health after controlling individually for other covariates ........................................... 118
Table B.8 Results of linear regression analyses of association between absolute income inequality and health, controlling individually for other covariates........................................................................ 119
Table B.9 Results of linear regression analyses in different models ..................................................... 120
Table B.10 Results of linear regression analysis of association between income inequality, absolute income and health within different levels of rural/urban status ........................................... 122
Chapter 1
Introduction

1.1 Background

Income and economic resources are known determinants of health (Sorlie et al., 1995). There are two major hypotheses surrounding the relationship between income and health: the absolute and relative income hypotheses. The absolute income hypothesis states that health is a linear function of income (Grossman, 1972). As members of communities have more income and subsequently more economic resources, they tend to eat better, they practice risky behaviors less often, they spend more money on health care, and therefore will be healthier (Veenstra, 2003). The relationship appears to be obvious in developing countries, but in the early 1990s Wilkinson showed that above a level of income of about $5000 per capita in aggregate (e.g., in a country), absolute standards of living have less impact on health and the role of the relative distribution of income becomes more prominent (Wilkinson, 1992, 1994).

Wilkinson’s hypothesis launched a great amount of debate in population health research and attracted the attention of social epidemiologists, public health professionals, and policy makers alike. Studies on relationships between income inequality and population health remain inconclusive. In countries with greater income imbalances where per capita income is high, but the relative income distribution is more skewed between the wealthy and the poor, the relative income hypothesis compared to absolute income hypothesis is supported more often. Also, when the units of analysis are large geographic regions, relative income appears to predict health status better when compared to absolute income (Wilkinson & Pickett, 2006). In contrast, in more egalitarian societies such as Canada and the Scandinavian countries, and smaller geographic regions such as counties, parishes, and neighborhoods the
studies tend either to negate the relative income hypothesis or to be inconclusive (Franzini et al., 2001; Wilkinson & Pickett, 2006). More than ten years after introducing the relative income hypothesis and despite dozens of papers published on this topic, the issue continues to be debated.

Is relative income inequality a real public health concern in Canada? The problem is more challenging in Canada as the few existing Canadian studies are either supportive of the absolute income hypothesis (Laporte & Ferguson, 2003; McLeod et al., 2003) or inconclusive (Sanmartin et al., 2003). Studies that include Canada have tended to use large geographic regions such as provinces or to examine health status in populations of one province or region. The hypothesis has not been tested across all Canadian health regions. This presented a unique research opportunity for the current thesis.

1.2 Objectives

The primary objective of this thesis was to test the “relative income” hypothesis (Wilkinson, 1992) that suggests that population health is a function of income distribution, independent of absolute income and other potential confounders, versus the “absolute income” hypothesis which implies that population health is a linear function of absolute wealth across Canadian health regions. Specific objectives were to:

1. Estimate changes in the relationship of interest after controlling for absolute income and other potential confounders.

2. Provide the best and the most parsimonious model that can explain variations in the health of the Canadian population.

A priori, we also wished to examine this theory in rural versus urban populations because of the interests of the CIHR program that funded this research.
1.3 Overview of the Study Design

To test the relative versus the absolute income hypothesis across Canadian health regions, an ecological design was employed. This design aimed to analyze the potential relationship between income inequality and self-rated health. The source of the data was the Canadian Community Health Survey (CCHS), 2005 cycle 3.1 (Statistics Canada, 2005). The units of analysis were health regions as defined by Statistics Canada. Because of the ecological nature of the analysis, all dependent, independent and potential confounding variables were measured at the group (health region) level.

The outcome variable (health status in populations) was a measure of self-rated health at the population level. It was estimated according to the percentage of people who rated their health as good or excellent in each Canadian health region. The primary exposure variable (relative income inequality) was the ratio of people in each health region whose income was less than $15,000 (CND) relative to those who made more than $80,000 in the year preceding the survey. This ratio provided a “range measure” of the distribution of income, and was based on an existing approach to the assessment of income inequality (Harper & Lynch, 2006). The main covariates that were controlled for in the regression analyses were absolute income, health care utilization, measures of socio-demographic status, risky behaviours such as smoking and alcohol consumption, rural/urban status of the region, and social capital.

Statistical analyses used in this thesis were descriptive, correlational and regression analyses. Simple and multiple linear regression models were employed to ascertain the amount of variance explained in the health status measure based on measures of relative and absolute income. These analyses controlled for potential confounders at the group level, and
established the best model to explain the relationship between independent and dependent variables.

1.4 Rationale

The main reason for conducting this study is its potential to influence policy. Evidence demonstrates that income inequality in Canada, after a period of stability across the 1980s, rose during the 1989-2004 period (Banting, 2005; Heisz, 2007). Measures of income inequality appear to be stable during the past two years (Statistics Canada, Table 202-0705). Meanwhile, Canadians have enjoyed a steady rise in their absolute income over the past decade. The median total income of Canadians increased (in 2006 constant dollars rates) from $21,000 in 1996 to $25,300 in 2006 and average income increased from $29,100 in 1996 to $34,400 in 2006 (Statistics Canada, 2008a, Table 202-0402). At the aggregate level, the percentage of people with total incomes of $30,000 and over has increased from 37 per cent to 43.7 per cent over the same time period (Statistics Canada, 2008b, Table 202-0402). Determination of whether absolute or relative income has an impact on the health of the Canadian population has important implications for health and social policies, and should be considered in economic policies related to income distribution.

Despite the long-time existence of Wilkinson’s seminal paper (Wilkinson, 1992) in which the “relative income hypothesis” was introduced, research on this topic within any one country remains scarce. Canada is a developed industrial country with high levels of economic development and as a country is well past the “epidemiological transition” (Wilkinson, 1994). This makes Canada a suitable place to study the relative income hypothesis. Few Canadian studies exist on this issue and no study so far has tested the relative income hypothesis across all Canadian health regions. By focusing on Canada as a single country, this thesis tests the competing hypotheses exclusively in Canada and can in part fill a
gap in knowledge. The research may shed light on health consequences of income inequality within Canada and provides additional evidence to support or reject the “relative income” hypothesis.

1.5 Thesis Organization

The second chapter of this thesis is a literature review that focuses on our current understanding of the relative and absolute income hypotheses as determinants of population health. Between and within country studies are presented separately, and Canadian studies are reviewed in more detail. The third chapter provides a detailed review of the methods used for measurement and analyses. Chapter 4 is an original manuscript to be published in a peer review journal. Chapter 5 is a discussion chapter that consists of the main findings and their interpretations, study limitations and strengths, along with the policy implications of the research and suggestions for directions for future research. The full study results are presented in appendix B.
1.6 References


Chapter 2
Literature Review

2.1 Literature Search Strategy

Studies of relationships between income inequality and population health were identified in Medline/Pubmed using the key words: ‘income inequality’, ‘population health’, ‘Canada’, ‘self-rated health’, ‘relative income hypothesis’ ‘absolute income hypothesis’, and ‘ecological analysis’. Reference lists of two systematic reviews (Wilkinson & Pickett, 2006) and (Lynch et al., 2004) were also reviewed for relevance and inclusion. Finally, to incorporate unpublished dissertations in the literature review, the Dissertation and Theses database in ProQuest (www.proquest.umi.com) was also searched using the same keywords, and relevant studies were included in the review.

2.2 Social Determinants of Health

The science of epidemiology seeks to identify risk factors for diseases. These risk factors may be an environmental agent such as a virus, a behavior such as smoking, or a change in the body’s biology such as an elevation in blood pressure. Social epidemiology, however, employs a different approach. Instead of focusing on individual risk factors, it seeks to identify whether societal factors influence both individual and population health and the mechanisms by which this occurs (Marmot & Wilkinson, 2006).

Among social determinants of health, one can list the following factors: social class (Turner & Marino, 1994), employment (Mathers & Schofield, 1998), type of job and the amount of control that individuals have in their job (Kuper & Marmot, 2003), social networks (Geckova et al., 2003), housing (Thomson et al., 2001), education (Raphael, 2003), and income (Sorlie et al., 1995). Despite their probable correlation, all of these factors can be
considered as independent social characteristics. For example, having a good social network may be studied independently of employment status (Marmot & Wilkinson, 2006).

Income is one of the most important social determinants of health, and has both direct and indirect effects. Income may influence housing, employment, health care expenditures, and other factors that impact upon health. Aspects of economy, namely absolute and relative income, are important determinants of health and social well-being (Lynch & Kaplan, 1997).

2.3 Income Inequality and Health

Relationships between material wealth and health appear to be obvious at both individual and aggregate levels. Higher income is a precondition for healthier environments, particularly in developing countries (Rodgers, 1979). As members of communities have more economic resources, they tend to eat better, they practice risky behaviors less often and they spend more money on health care (Veenstra, 2003). A community with more healthy people will report a better health status. If countries are divided into two clusters, (i.e., poor and rich) and the health of the two clusters is compared, richer countries exhibit better health. This is the concept surrounding the “absolute income” hypothesis (Grossman, 1972) and there is little debate about the potential benefits that higher income have on improving individual health.

Another hypothesis that has gained popularity in recent years is the “relative income” hypothesis (Kawachi et al., 1999) which asks, “within an aggregate population, does distribution of wealth have an additional impact on individual health?” This implies that health is a function of relative wealth, controlling for absolute wealth and other health-related factors.

2.3.1 The Absolute Income Hypothesis

The absolute income hypothesis suggests that the health of an individual depends on their absolute level of income. In other words, this hypothesis states that it is income level that
matters for health, not income relative to other people. One of the pioneers of this hypothesis was Grossman who in an economic theory in 1972 said that an increase in income will increase investments in health-enhancing goods (Grossman, 1972). In 1979, Rodgers found a non-linear concave relationship between income and measures of population health including life expectancy and infant mortality rate at international levels. This suggests that an increase in absolute income will not improve health after a certain amount of economic development (Rodgers, 1979).

Numerous country-level studies also support the absolute income hypothesis. For example, Pritchett and Summers (1996) in a cross-country time series analysis found a positive relationship between gross domestic product (GDP) per capita of countries as measures of absolute income, and life expectancy and infant mortality rates as measures of population health. Similar results have been found by other investigators (Preston, 1975; Pritchett & Summers, 1996; Rodgers, 1979). In another country level study, Lutter and Morrall (1994) have shown that willingness to pay for health services is positively related to level of income independent of relative income. A recent Canadian study using the 1998-99 Canadian National Population Health Survey data, provided evidence in support of the “absolute income hypothesis” (Stengos & Sun, 2006). Nevertheless, the same studies that support the “absolute income” hypothesis (Preston, 1975; Rodgers, 1979) found that above a certain threshold, gains in life expectancy were not related to higher levels of income (i.e. the relationship was non-linear).

### 2.3.2 The Relative Income Hypothesis

During the 1970s and 1980s, there were sporadic reports that health may be linked to distribution of income (Legrand, 1987; Pampel & Zimmer, 1989). No particular hypothesis was introduced until Wilkinson’s seminal article in 1992 (Wilkinson, 1992)
that brought the “relative income” hypothesis into the spotlight. Wilkinson, in a study of life expectancy and gross domestic product per capita using data from 1970 and 1990, showed that “there is some minimum level of income (around $5,000 per capita in 1990) above which the absolute standard of living ceases to have much impact on health” (Wilkinson, 1994, p. 62). In the process of economic development, there is a point at which the vast majority of the population gains reliable access to the basic material necessities of life. This point matches the “epidemiological transition” (Wilkinson, 1994); the shift in the main causes of death from infectious diseases to degenerative cardiovascular diseases and cancers. After this transition, health within developed countries shows a relationship with measures of socioeconomic status and is affected by relative income (Wilkinson, 1994).

This hypothesis triggered a large amount of debate and prompted numerous studies with conflicting results. Some studies, particularly by analyzing data for regions within the United States (Wilkinson, 2005), demonstrate strong association. International studies and studies from more egalitarian countries such as Sweden and Canada show non-conclusive results (Wilkinson & Pickett-, 2006). Results of studies on the topic remain inconsistent. The inconsistency among findings may lie in the heterogeneities of the study designs and statistical methodologies employed (Wen et al., 2003). Apart from some studies that employ path analysis (Gold et al., 2002), little work has been done to develop a design framework to investigate the link between income inequality and health. To date, the relative income hypothesis is neither refuted nor confirmed.

Different studies at different geographic levels will be reviewed in more detail in section 2.4. Since the objective of this thesis is to examine this relationship in Canadian
health regions, more emphasis will be put on the Canadian studies and those based on smaller regions.

2.3.3 Measures of Income Inequality

Income is not distributed equally in society. At a population level, some enjoy larger amounts of income whereas some experience deprivation. Measurement of income inequality is one of the major concerns of economists and policy makers (Cowell, 1995). Although income is a characteristic of an individual, income inequality is not reducible to characteristics of individuals (Szklo & Nieto, 2007). Societies have income distributions whereas individuals do not (Wilkinson, 1994).

There is no “gold standard” measurement for income inequality (Kawachi, 1999, p. xviii). One conventional way is to cite a range that compares two extreme categories and reports a ratio (e.g., the ratio of the number of poorest people to richest in an aggregate population). These measures are simple to calculate and interpret, do not need individual data, but ignore middle income categories and relative population sizes (Harper & Lynch, 2006). Share ratios using population income decile(s) such as the bottom or top percentage share (Waldmann, 1992; Wilkinson, 1992), and 90th/10th percentile (Lynch et al., 1998) or other share ratios (Daly et al., 1998) are the common range ratios used in the study of relationships between income inequality and health status in populations. Most census data sources provide users with the share of total income for each decile of population from poorest to richest. These data can be used in different manners, such as the calculation of share ratios of deciles or the “bottom 70 per cent share” as used by Wilkinson (1992) or the “top 5 per cent share” as used by Waldmann (1992).

Other measures used in studies on income inequality include disproportionality measures such as the Gini coefficient, the Theil’s index, and the Mean log deviation (Harper & Lynch,
These indicators consider population size of each category, and account for the middle income groups. These measures are more difficult to calculate and interpret in population health research, and also rely upon individual level data that are not always available.

The basic concept is to measure the imbalance between share of population and share of the variable (ill health, income). The most commonly used disproportionality indicator is the Gini coefficient (De Maio, 2008; McLaughlin & Stokes, 2002; Olsder et al., 2003; Wilkinson, 1992). Other measures such as the Robin Hood index (Kennedy et al., 1996; Walberg et al., 1998), Atkinson Deprivation index (Laporte, 2002), Theil Entropy Index (Lynch et al., 1998) have also been employed.

2.3.4 Measures of Health

Health is a complex outcome that can be measured across a continuum of general to specific indicators. There is no consensus on how to quantify “health” and various authors have used different proxy measures. A review of studies of relationships between population health and income distribution indicate the use of two main types of indicators: 1) objective indicators such as life expectancy or mortality rates, and 2) subjective indicators such as self-reported health.

Life expectancy is one of the key indicators of population health and development and has been used by Wilkinson (1992). Mortality rates have been used extensively as indicators for health at a national level (Ben-Shlomo, 1996; Lynch et al., 1998). Infant mortality has been used mainly in studies at an international level and in developing countries (Casas et al., 1998; Hales et al., 1999; Waldman, 1992).

Self-rated health is a subjective indicator of health that has been used in most recent studies (Blakely et al., 2002; Subramanian et al., 2001). People answer a simple question,
such as that used in the CCHS: In general, would you say “YOUR” health is? 1- Excellent 2- Very good 3- Good 4- Fair 5- Poor. There is some evidence suggesting poor validity (Sturm & Gresenz, 2002) and reliability (Crossley & Kennedy, 2002), of this general measure of health status. In an Australian study (Crossley & Kennedy, 2002), a subset sample of respondents was asked the self-reported health status question twice; before and after an additional set of health-related questions. Among participants approximately 28 per cent changed their response; 13.6 per cent reported a higher level of health whilst 14.8 per cent reported a lower level. In contrast, Idler and Benyamin (1997), in a review of 27 studies in United States and international journals, showed that self-rated health is an independent and accurate predictor of mortality in nearly all studies. This was true despite variation in specific health status indicators and other relevant covariates such as behavioural or psycho-social factors known to affect mortality. In all 27 studies, negative evaluations of health were associated with higher relative odds of mortality. In a previous study, the same author showed that self-ratings of health in 1980 were associated with changes in functional ability over periods of one through six years (Idler & Kasl, 1995).

2.4 Potential Causal Mechanisms

The ways that income inequality affect health are not clear. Most research on this topic is observational and ecological, although a few cohort studies exist (Fiscella & Franks, 1997; McLeod, 2003). Since no trial has been done or is feasible to perform at population levels to test this relationship, it is difficult to provide high levels of epidemiological evidence and suggest an evidence-based causal mechanism. In theory, it is reasonable to think of five potential causal mechanisms: 1) an individual income mechanism; 2) a neo-material mechanism; 3) a social capital mechanism; 4) a psycho-social mechanism, and 5) a socio-
biological mechanism. Potential causal mechanisms by which income inequality could result in poor health are shown in figure 2.1 in the form of a conceptual map.

![Theoretical Causal Mechanisms](image)

**Figure 2.1: Theoretical Causal Mechanisms**

### 2.4.1 Individual Income Mechanisms

The individual income mechanism indicates that observed aggregate level associations between income inequality and health may simply represent inadequately measured rates of income differences at the individual level (Fiscella & Franks, 1997). This interpretation is analogous to the “absolute income” hypothesis, which suggests that health status in populations is merely a sum of the individuals’ health within that population (Diez-Roux, 1998; Koopman & Lynch, 1999).
2.4.2 Neo-material Mechanisms

The concept of a neo-material mechanism was first suggested by Lynch in response to the “relative income” hypothesis. He stated that poor health in countries with unequal distributions of income is attributed to the direct effects of a lack of material facilities (Lynch et al., 2000). In this view, a society with greater income inequality will have a higher percentage of people with low incomes, and this higher prevalence of poor people accounts for the relation with poor health (Lynch et al., 2000). A lower income limits an individuals’ access to material resources such as education, employment, health care, and proper housing, and subsequently affects their health (Kaplan et al., 1996).

2.4.3 Social Capital Mechanisms

According to the Last’s Dictionary of Public Health, social capital is defined as “the processes and conditions among people and organizations that lead to their accomplishing a goal of mutual social benefit, usually characterized by interrelated constructs of trust, cooperation, civic engagement, and reciprocity, reinforced by networking” (Last, 2007).

Social capital can be linked to the health of individuals and communities (Edmondson, 2003). Theoretically, widening the gap between the poor and rich might result in damage to the social fabric of a community through segregation of people into poor and rich residential regions, hence decreasing the social capital of the community and consequently having ill effects on population health (Kawachi et al., 1997). Evidence from a prospective cohort study (Kawachi et al., 1996) shows that isolated people are at increased risk of cardiovascular disease mortality (RR=1.9; 95% CI=1.07-3.37) as well as stroke (RR=2.21; 95% CI=1.12-4.35). Kawachi et al. (1997) in a study of 39 American states showed that both the degree of civic distrust and paucity of other measures of social capital (group membership, black-white equality) were strongly correlated with overall mortality. In their ecological regression model,
variation in level of social trust explained 58 per cent on observed the variance in total mortality.

2.4.4 Psycho-social Mechanisms

Individuals living in an unequal society do not have a good social support system and lack total control over their life. This situation creates a sense of hopelessness and insecurity. Some studies have shown that hopelessness (Everson, 1997) and job insecurity (Bosma et al., 1997) are attributed to poor health. These situations can affect health directly through stress-induced behaviours such as alcohol and drug abuse, smoking, or indirectly through chronic stress (Berkman, 1995; Kawachi et al., 1999). Other effects of inequality are negative emotions such as distrust and shame due to perception of place in the social hierarchy based on relative position according to income (Wilkinson, 1997b). These emotions can be translated into antisocial behaviours, less cohesion within societies, and less civic participation. The end result is poor health status in populations.

2.4.5 Socio-biological Mechanisms

In studies on non-human primates, low social status is associated with a worse cholesterol profile, increased atherosclerosis, more obesity, and depression (Shively 1994, 1997). It may not be reasonable to generalize a physiological experiment in primates to human populations; however, Brunner (1997) found that more infection and cardiovascular disease were observed in disadvantaged groups of human society.

There are some theories that argue how social experience can get “inside the body”. The effect of social support on immunological and neuro-endocrine systems is well documented. Stress can weaken immune function (Kennedy et al., 1990) especially cellular immunity (Schleifer et al., 1983) and the neuro-endocrine system (Seeman & Robins, 1994). These studies suggest that income inequality, by creating stress may result in maladaptive reactions
in the immune and neuro-endocrine systems. Findings of an ecological study of health differences between two cities in Sweden and Lithuania also provide evidence that chronic stress due to social gradients will have a direct ill effect on the endocrine system (Kristenson, 1998).

2.4.6 Alternative Explanations

There are other explanations for the observed associations between income inequality and health. The most controversial one is the “evolutionary psychological” theory. This theory implies that all hazards to health in contemporary societies are evolutionarily novel, and more intelligent people are better able to cope with these hazards and hence are healthier. Kanazawa (2006) in a study on 126 countries shows that income inequality and economic development have no effect on life expectancy at birth, infant mortality and age-specific mortality, net of average national intelligence quotient (IQ). He suggests that individuals in wealthier and more egalitarian societies live longer and stay healthier, not because they are wealthier or more egalitarian but because they are more intelligent. This theory has many opponents. Wilkinson dismisses this theory and believes that the role of societal factors on intelligence should not be ignored (Wilkinson, 2007) and Dickins et al. (2006) question the methods that were used in this study with regard to sample size and sampling, extrapolation, and inconsistency across measures.

2.5 Review of Studies on the Relationship between Income Inequality and Health

2.5.1 International Studies

International studies generally demonstrate a link between income inequality and population health. There is some evidence that the relationship will vary by population size and the hypothesis will be more strongly supportive in larger geographic regions (Franzini et
al., 2001). This is one plausible explanation for more positive results in international studies, which compare large geographic regions. Research on the topic has been conducted within both developed and developing countries. Since the socioeconomic status and health infrastructure of these two groups of countries are very different, the relationship of interest may show different manifestations, and hence are reviewed separately.

2.5.1.1 Developed Countries

The “relative income” hypothesis was introduced in 1992 by Wilkinson after studying the relationship between income inequality and life expectancy in 12 industrial European countries, the United States and Canada (Wilkinson, 1992). Before Wilkinson’s seminal article there were a few studies in the late 1980s, in which a positive relationship between income inequality and mortality rates was found (Legrand, 1987; Pampel & Pillai, 1989). Subsequent international studies in the following years, using different measures of mortality rates, were also able to demonstrate a link between income inequality and measures of poor health status in populations of developed countries (Duleep, 1995; Macinto et al., 2004; McIsaac & Wilkinson, 1997). However, the results were not conclusive. In some studies the relationship disappeared after controlling for covariates such as Gross Domestic Product (Mellor & Milyo, 2001), deprivation (Bobak et al., 2000), and poverty (Judge, 1995). Some other authors found no relationship between an aggregate level measure of health and income inequality (Lynch et al., 2001). International studies generally compared various mortality rates (Judge, 1995) or life expectancy (Wilkinson, 1992) across different countries. Infant mortality rates were usually used in developing countries, although Wennemo (1993) used this health measure and showed that higher income inequality was significantly associated with higher infant mortality rates in 18 industrial countries (Wennemo, 1993). Two recent theses tested the hypothesis at an international level; Cho (2004) in a study of 100 developed
and developing countries showed that even after accounting for absolute income, income inequality had a statistically significant contribution to the health status of populations. In a PhD dissertation by Goyal (2003), the relative income hypothesis did not hold when longitudinal data from fifteen developed countries was used.

In summary, international studies from developed countries generally show positive results. This may be due to two demonstrated facts: 1) the relationship of interest tends to be positive in larger geographic regions (Frazini et al., 2001; Wilkinson, 1997), and 2) the “relative income” hypothesis appears to be valid in countries that have passed the “epidemiological transition” phase (Wilkinson, 1994), which is the case in most developed industrial countries.

**2.5.1.2 Developing Countries**

Even though Wilkinson concluded that the “relative income” hypothesis is valid for the countries that have passed the “epidemiological transition” phase (Wilkinson, 1994, 1992), examination of this hypothesis within and across developing countries also provides positive results (Drain et al., 2004; Hales et al., 1999; Waldman, 1992). Earlier studies were mainly focused upon infant mortality as a measure of population health (Flegg, 1982; Waldman, 1992) and all supported the validity of this theory in developing countries. Waldman (1992) found that after controlling for gross domestic product per capita, infant mortality rates were higher for countries with unequal wealth distribution. Hales et al. (1999) found similar trends and showed at all levels of economic development that infant mortality rates tend to be lower in more egalitarian societies.

In contrast, Mellor et al. (2001) showed that after longitudinally re-examining the relationship across thirty countries over five decades, the relationship disappeared. In another
study controlling for some covariates and using individual health data, no relationship has found (Beckfield, 2004).

### 2.5.2 Within Country Studies

Studies conducted at national levels show different results. Results of studies situated within larger and more unequal places such as the United States are mostly positive, whereas in smaller and more egalitarian regions such as European countries the relative income hypothesis has not been supported.

#### 2.5.2.1 United States Studies

Several earlier studies from the late 1980s and early 1990s showed positive relationships between income inequality and various measures of mortality across American states (Kaplan et al., 1996; Kennedy, Kawachi, & Prothrow-Stith, 1996; Kawachi & Kennedy, 1997; Shi et al., 1999). More recent studies using multilevel analyses (Kahn, 2000; Subramanian & Kawachi, 2003a) also found results in support of the relative income hypothesis in American states. However, in some studies, after controlling for socio-demographic factors such as percentage of black population (Mellor & Milyo, 2001), and education (Muller, 2002), the relationship disappeared.

Mortality rates and self-rated health are not the only health outcome measures used in studies of the relationship between income inequality and health in the United States. Holtgrave and Crosby (2003) studied the association between income inequality and infectious diseases, namely sexually transmitted infections (STIs), and found a positive relationship. In 2004, the same investigators tested this theory on rates of tuberculosis and found similar results (Holtgrave & Crosby 2004). Health outcomes that have been studied less frequently in the United States include stroke mortality (Shi et al., 2003), the teen birth rate and violence (Pickett, Mookherjee, & Wilkinson, 2005). In addition, in a number of
American studies homicide was found to be positively associated with income inequality (Bailey, 1984; Balkwell, 1990; Daly et al., 2001).

Not all US studies support the relative income hypothesis (Mellor & Milyo, 2002; Muntaner et al., 2004). Sturm and Gresenz (2002) in a national survey of 60 randomly selected regions of the United States found no association between income inequality and 17 different chronic physical and four psychiatric disorders. They stated that one possible reason for previous positive relationships is using an inappropriate health measure (i.e. self-rated health) (Sturm & Gresenz, 2002).

2.5.2.2 Other Developed Countries

Research in developed countries other than the United States mostly failed to identify a relationship between income inequality and health (Wilkinson & Pickett, 2006). However, there are a few supportive studies in Italy (De Volgi et al., 2005), Russia (Walberg et al., 1998), and the United Kingdom (Stanistreet et al., 1999). In a Japanese study, after controlling for mean area income, income inequality was not associated with poor self-rated health (Shibuya et al., 2002). The same results was observed when individual level characteristics such as age, ethnicity, rurality, household income, and regional mean income, were controlled for in a New Zealand analysis (Blakely et al., 2003). A longitudinal time-series analysis in Israel showed that after controlling for gross domestic product per capita, inequality in income did not affect population health over time. However, reductions in inequality over time were associated with better population health (Shmueli, 2004).

One study conducted in the United Kingdom by Weich et al. in 2002 found limited evidence of an independent association between regional income inequality and poorer self-rated health. The association seemed greater among those with the lowest income. Similar trends were observed when common mental disorders were examined as health outcomes.
(Weich et al., 2001). Two studies in Spain failed to show any relationship between relative income and prevalence of long term disability (Regidor et al., 1997) and life expectancy (Regidor et al., 2003).

In a study in Denmark, income inequality was not associated with age standardized mortality after adjustment for individual risk factors, whereas individual household income was (Osler et al., 2002). The same authors in 2003 found a weak association between income inequality and ischemic heart disease in Danish men and women (Osler et al., 2003). As expected, no effect was observed in the studies of countries with smaller imbalances in income such as Finland (Blomgren et al., 2004), Belgium (Lorant et al., 2001), and the Netherlands (Drukker et al., 2004) where the levels of income inequality are low (Asada, 2007).

### 2.5.2.3 Canadian Studies

Most existing studies conducted in Canada support the “absolute income hypothesis”, even though the objectives of most of them have been to test the “relative income hypothesis”. One can divide these studies into two groups: 1) mixed United States and Canadian studies 2) exclusively Canadian studies.

In 2000, Ross included data for the Canadian provinces as well as the fifty States and showed that income inequality is strongly associated with mortality in the United States and in North America as a whole (Ross et al., 2000), but there is no relation within Canada at either provincial or metropolitan area levels (Ross et al., 2000); findings were confirmed at the provincial level by Laporte and Ferguson (2003), at the metropolitan level by McLeod et al. (2003), and at the neighborhood level by Hou and Chen (2003). However, a study on homicide rates in 50 American States and 10 Canadian provinces showed that after controlling for median household income, higher income inequality was associated with
higher homicide rates in both the United States and Canada (Daly et al., 2001). One study on British Columbia’s coastal communities also demonstrated that after controlling for mean household income there is no relationship between income inequality and age-standardized mortality rates (Veenstra, 2002b). Humphries and Doorslaer (2000) research provided more support for the absolute income hypothesis in Canada. They showed that the higher the level of income, the better the level of self-assessed health in the Canadian population (Humphries & Doorslaer, 2000).

Some Canadian studies show inconclusive results. A study on Toronto neighborhoods (Hou & Myles, 2005) reported that after individual low-income status was taken into account, neighborhood income inequalities were not associated with individuals’ reported numbers of chronic conditions. However, income inequality at the neighborhood level remained significantly associated with poor self-rated health. Another study claims that using inequality in the distribution of market income (income before the deduction of taxes or addition of benefits) rather than total or after tax income seems to explain why some Canadian cities are healthier than others (Sanmartin et al., 2003). Additional studies have shown positive relationship between income inequality and death rates in Saskatchewan (Veenstra, 2002a), and homicide in Canadian urban regions (Kennedy et al., 1991).

Finally two dissertations on this topic based on Ontario public health units showed contradictory results. In Phillips’ (2002) ecological research, after controlling for median income, the contribution of income inequality on health was not significant anymore, whereas a multilevel study by Xi (2003) indicated a statistically significant association between income inequality and health independent of individual income.
2.5.2.4 Developing Countries

Within country studies in the developing world are more likely to find a positive relationship between income inequality and population health. The majority of the literature comes from South American countries where a positive relationship between infant mortality rates and income inequality has been reported (Casas & Dachs, 1998). Subramanian and Kawachi (2003b) in a study of about 100,000 adults in 285 communities of Chile found that higher inequality in income is associated with 20 per cent increased odds of reporting poor self-rate health. One study in Brazil was supportive of the relative income hypothesis; Szwarcwald et al., in 1999 found that higher income inequality is associated with higher homicide rates within administrative regions of the city of Rio de Janeiro (Szwarcwald et al., 1999).

In two other Brazilian studies the association between population health and income inequality disappeared after adjustment for illiteracy (Messias, 2003) and poverty (Szwarcwald et al., 2002). In Ecuador, where chronic malnutrition still affects 26 per cent of children under five, after controlling for relevant covariates, economic inequalities at the provincial level had a statistically significant association with chronic malnutrition; however, at municipal or local levels, inequality was not associated with this health outcome (Larrea & Kawachi, 2005).

In an attempt to examine the robustness of the income inequality–population health relationship in Argentina, De Maio (2008) used five different income inequality indices (each sensitive to inequalities in different parts of the income spectrum) in relation to five measures of population health. The study found a correlation between life expectancy and income inequality, but this association was not consistent across all five income inequality indices. Three out of five measures of population health, (infant mortality, self-reported poor health,
and self-reported activity limitation) were not correlated with any of the income inequality indicators.

The only study found outside South America in a developing country was performed in Taiwan. The study explored the effect of income inequality on mortality for three years (1976, 1985, and 1995) with positive results for all of them. Stronger associations were found in 1995. The authors concluded that the health of the population is affected more by relative income than by absolute income after a country has changed from a developing to a developed economy (Chiang, 1999).

### 2.6 Effect of Size of Regions on the Relationship of Interest

Compared to studies conducted in large geographic regions such as countries and states, studies of smaller regions such as counties, parishes, or neighborhoods are more likely to show no relationship between inequality in income and health (Fiscella & Franks, 1997; Hou & Myles, 2005; Jesmin, 2004; Veenstra, 2002b). These findings suggest that the size of community affects the relationship between income inequality and health. A review of 168 studies showed that 83 per cent of international studies were totally supportive of the hypothesis. In the large sub-national geographic regions such as states and provinces this ratio declined to 73 percent. In small geographic regions, only 45 per cent of studies supported the relative income hypothesis (Wilkinson & Pickett, 2006). Wilkinson argues that the impact of income inequality on health comes from social stratification and this factor is lost in small, socially homogeneous regions (Wilkinson, 1997a). In small regions, there is no social heterogeneity, hence no social stratification. Therefore, the health status of populations is a function of absolute income not relative income (Wilkinson, 1997b). In moving from larger to smaller regions, median income becomes a more important predictor, and income inequality a weaker predictor of mortality (Wilkinson, 2000).
Franzini et al. tested this theory on Texas counties and found that in counties with a population greater than 150,000, mortality is positively related to income inequality, whereas among counties with the population less than 150,000, median income affects mortality (Franzini et al., 2001). Jesmin (2004) found that income inequality of a county is not a direct predictor of infant mortality rates in Texas counties.

2.7 Effect of Rural/urban Status on the Relationship of Interest

Few studies have controlled for urban/rural status in the examination of relationships between income inequality and health. In three existing studies, after controlling for rural/urban status and other factors, income inequality was not associated with alcohol dependency (Henderson et al., 2004), infant mortality rate, all cause mortality, the cause-specific mortality rate (Mellor & Milyo, 2001), and self-rated health (Subramanian et al., 2001). In contrast, other studies that controlled for the percentage of the population living in an urban area did not change the positive relationship between income inequality and shorter heights (Steckel, 1983), the infant mortality rate (Waldman, 1992), and rates of homicide (Kennedy et al., 1998).

2.8 Methodological Issues

A number of methodological issues are important to consider surrounding the relative income hypothesis. Gravelle (1998) interprets the result as a statistical artifact, supported by the facts that 1) the relationship between wealth and health is not linear and 2) the use of aggregate rather than individual data creates this artifact as an example of the "ecological fallacy". Some researchers argue that the existence of the relationship depends on the choice of income inequality measure used (Judge, 1995). Still others indicate that the relationship will disappear after properly controlling for individual potential confounders such as absolute income (Fiscella & Franks, 1997). There is also some debate on the nature of the relationship
between income inequality and health. It is speculated that this relationship may depend on the level of inequality and be apparent only above a certain income inequality threshold (Asada, 2007).

2.8.1 Artifactual Relationship between Income Distribution and Health

Some argue that the observed relationship between income inequality and health is in fact artifactual. The curvilinearity of relationships between income and health at the individual level creates “artifactual” associations between income inequality and health at the population level (Ellison, 2002; Gravelle, 1998; Wildman 2001).

2.8.2 Ecological Fallacy/Atomistic Fallacy

Typically, existing analyses are either based on aggregate or individual levels (Sanmartin et al., 2003), yet the majority of the research on the relationship of interest is ecological. The biggest methodological challenge to ecological studies, whose units of analysis are an aggregate level, is the concept of the “ecological fallacy” (Greenland & Robins, 1994). Last’s dictionary of epidemiology defines this concept as “the bias that may occur because an association observed between variables at an aggregate level does not necessarily represent the association that exists at an individual level.” (Last, 2001, p.56). Some authors argue that it would be an ecological fallacy to infer that observed relationship at the population level necessarily means that individual health is also influenced by income distribution (Gravelle, 1998). Individual analyses, on the other hand, ignore the socio-economic context within which individuals experience differential levels of health and are prone to the “atomistic fallacy” (Macintyre, 2000). The latter is defined as an “erroneous inference about group or ecological relationship made on the basis of associations observed at the individual level” (Last, 2001, p. 8). This can be a methodological problem for research on the relationship of interest when income and health are measured at the individual level and inference is made at
the ecological level. If a study only utilizes single-level measurement (aggregate or individual) and avoids making any cross-level inferences, none of these fallacies should be an issue (Susser, 1994a, 1994b).

2.8.3 Multi-level Approach

In responding to methodological concerns mentioned in 2.8.2, multi-level approaches are one potential solution. Some argue that relationships between income inequality and health are intrinsically multilevel (Subramanian et al., 2001). Applying ‘single level’ analyses either at aggregate or individual levels cannot explain why more equal regions are healthier, because this approach is unable to differentiate compositional effects (what is in a place) from contextual factors (the difference a place makes) (Subramanian et al., 2001). By using the single level approach, researchers are unable to verify that better health in more egalitarian societies are due to a relative lack of poor people or are a direct effect of inequality itself (Wilkinson, 2005).

2.8.4 Choice of the Income Inequality Measure

Different inequality measures may reflect somewhat different income distribution patterns (Bobak et al., 2001). The use of different measures has raised two questions: 1) which measure is able to assess the income inequality more accurately and validly? 2) can using different measures yield different results? There are no definitive answers to these questions. Some studies have used more than one measure for inequality and the relationship between income inequality and health was not the same among the two measures. For example, Kennedy et al., (1996) used both the Robin Hood index and Gini coefficient to measure income inequality. They showed that the Robin Hood index was positively correlated with total mortality and causes of death controlling for age and poverty, while the Gini coefficient showed very little correlation with any of the causes of death (Kennedy et al., 1996).
However, Kawachi and Kennedy (1997) showed that the choice of measure of inequality does not appear to affect the relationship with mortality, and measures are typically highly correlated with each other ($r > 0.8$).

2.9 Summary

In short, based on the identified studies, studies on relationships between income inequality and population health are not conclusive. There is a trend towards more positive results in less equal countries such as the United States and the United Kingdom, and more negative results in more egalitarian societies such as Canada and the Scandinavian countries. Another observed trend is that more positive results are reported from large geographic regions such as countries, states, and provinces and less positive results from smaller geographic regions such as counties, parishes, or neighborhoods. Existing Canadian studies are either supportive of the absolute income hypothesis or inconclusive. There is no study based on all Canadian health regions. The objective of this research is to perform such a study.
2-10 References


Chapter 3
Methods

3.1 Overview

This thesis involved an ecological analysis of the potential relationship between income inequality and self-rated health across Canadian Health Regions. The data source was the Canadian Community Health Survey, CCHS 2005 cycle 3.1 (Statistics Canada, 2005a).

3.2 Purpose and Empirical Objectives

The purpose of the thesis was to develop new knowledge surrounding how the unequal distribution of income affects population health. This was addressed through consideration of “self-rated health” as a proxy measurement for population health (primary outcome) and its relationship with an indicator of income inequality (primary exposure). The thesis was developed in manuscript format. Study objectives were as follows:

Objective 1: Inequality and health, comparison between Canadian Health Regions

This objective tests the ‘relative income’ hypothesis (Wilkinson, 1992) that suggests that population health is a function of income distribution, independent of absolute income and other potential confounders.

Objective 2: Inequality and health, comparison across rural/urban regions

Objective 2 tests the ‘relative income’ hypothesis that health is a function of income distribution, and further examines this hypothesis in rural versus urban populations.
3.3 Ecological Model

The research aimed to evaluate the effect of income inequality on population health. The units of analysis were Canadian Health Regions as described and grouped by the CCHS. Since all variables were measured at an aggregate level, an ecological model was an appropriate approach to apply to this study (Szklo & Nieto, 2007).

Susser (1994) describes four types of ecological studies: 1) obligate and apt, 2) optional but apt, 3) optional, not apt but convenient, and 4) neither obligate, apt, nor convenient. Ecological studies are obligate when they are the only choice available, either because of the research question (examining differences between groups) or because of the nature of a variable (variables that only apply to the group level such as income inequality) or merely because of the lack of individual data. Income distribution is a characteristic of a social system and it is not measurable in individuals (Lynch, 1997). Based on Wilkinson’s argument: “as relative income is inherently a social concept it can not be dealt with at an individual level: societies, not individuals, have income distributions.” (Wilkinson, 1994, p. 66), any study on income inequality, falls into the category of “obligate” ecological studies because the only choice for measuring income inequality is an ecological measure. A major limitation of ecological studies is their inability to control for cross-level confounding between contextual exposures (such as income distribution) and individual level exposures (e.g. personal income) and health outcomes (Blakely et al., 2002). One remedy for this problem is the application of multi-level analyses. By integrating a group level and individual levels into ecological studies multiple levels of determinants can be examined simultaneously. Similar to other ecological studies, our purely ecologic variables (social capital, income inequality, prevalence of smoking, etc) were related to purely ecological.
outcomes (percentage of people with at least good self-rated health) in each health region and findings do not apply to predictions about individual health characteristics (Schwartz, 1994).

There were two reasons for choosing health regions as units of analysis in this ecological study: 1) since there is no other similar study for Canadian health regions, this research can, in part, fill a gap in knowledge; and 2) people compare themselves with their near equals (Runciman, 1966) and near equals live close to each other. Some believe that small regions are better suited for studying the relationship between income inequality and health. In small regions, people have more opportunity to compare themselves to individuals who live in close proximity; hence, the measurement of the effects of social comparisons on the health status in populations is more feasible in such regions (Wilkinson & Pickett, 2006).

3.4 Source of Data

The Canadian Community Health Survey (CCHS) Cycle 3.1 is a national level survey of about 130,000 persons conducted in 2005. The target population included household residents aged 12 and older in all provinces and territories. Populations on Indian Reserves or Crown lands, those residing in institutions, full-time members of the Canadian Forces Bases, and those in some remote regions were excluded.

The primary objective of CCHS Cycle 3.1 and previous cycles of CCHS is to provide timely cross-sectional estimates of health determinants and to gather health related data at sub-provincial levels of geography (Health Regions) (Statistics Canada, 2005a). The general aim of providing CCHS data is to increase understanding of the relationship between health status and health care utilization and to aid in the development of public policy.

The survey is conducted by both personal and telephone interviews using computer-assisted interviewing software. The CCHS Cycle 3.1 questionnaire consists of three parts: common content, which was asked of all respondents; optional content, in which regions
asked selected questions; and sub-sample content, in which to ease response burden three sets of questionnaire were asked only of a subset of respondents (Statistics Canada, 2005a). The CCHS cycle 3.1 employed various strategies to obtain a high response rate such as introductory letters, multiple contacts at different times, and refusal conversion methods. CCHS achieved a household-level response rate of 84.9 per cent, a person-level response rate of 92.9 per cent, and a combined national response rate of 79 per cent. The dependent variable, the independent variable, and all covariates were measure at a group level and were obtained from the CCHS cycle 3.1 data (A list and brief descriptions of all variables are summarized in table 3.1).
<table>
<thead>
<tr>
<th>Name in CCHS</th>
<th>Label</th>
<th>Concept</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENEDHDI</td>
<td>Health</td>
<td>Population Health</td>
<td>Percentage of people with good, very good and excellent health in each health region.</td>
</tr>
<tr>
<td>INCEGPER</td>
<td>Inequality</td>
<td>Income Inequality</td>
<td>Ratio of number of people whose income is less than $15,000 to those who make more than $80,000 in each health region.</td>
</tr>
<tr>
<td>INCEGPER</td>
<td>Absolute</td>
<td>Absolute Income</td>
<td>Percentage of people whose income is more than $30,000 in each health region.</td>
</tr>
<tr>
<td>HCUE_1AA</td>
<td>MD</td>
<td>Health Care Utilization</td>
<td>Percentage of people who have a regular medical doctor in each health region.</td>
</tr>
<tr>
<td>DHHE_Sex</td>
<td>Sex</td>
<td>Sex</td>
<td>Percentage of males in each health region.</td>
</tr>
<tr>
<td>DHHEGAGE</td>
<td>Age</td>
<td>Age</td>
<td>Percentage of people who are more than 65 years old in each health region.</td>
</tr>
<tr>
<td>EDUEDR04</td>
<td>Education</td>
<td>Education</td>
<td>Percentage of people who have at least some post secondary education in each health region.</td>
</tr>
<tr>
<td>SMKEDSTY</td>
<td>Smoking</td>
<td>Daily Smoking</td>
<td>Percentage of people who are daily smokers in each health region.</td>
</tr>
<tr>
<td>ALCE_6</td>
<td>Alcohol</td>
<td>Alcohol Consumption</td>
<td>Percentage of people who consume more that 12 drinks regularly each week in each health region.</td>
</tr>
<tr>
<td>GEOOnDUR2</td>
<td>Rural</td>
<td>Rural /Urban Status</td>
<td>Percentage of people who live in rural regions in each health region.</td>
</tr>
<tr>
<td>GENE_10</td>
<td>Belonging</td>
<td>Social Capital</td>
<td>Percentage of people who rate their sense of belonging to their local community as strong in each health region.</td>
</tr>
</tbody>
</table>

Table 3.1: Description of Dependent and Independent Variables

3.4.1 Features of the CCHS cycle 3.1

Some special features of the CCHS cycle 3.1 are as follows:

3.4.1.1 Design

The CCHS cycle 3.1 sampling plan is a multistage stratified cluster design in which the dwelling is the final sampling unit. In the first stage, homogeneous strata of rural/urban status and socio-economic characteristics are formed, and independent samples of clusters are drawn from each stratum. In each stratum, six clusters are chosen by a random sampling
method. The size of each cluster corresponds to the number of households. In the second stage, within each cluster, dwellings or households are selected by using a systematic sampling method (Statistics Canada, 2005a).

3.4.1.2 Weighting

To derive meaningful estimates, the results need to be weighted (i.e., each person included in the sample is considered to "represent" many other persons not included in the sample) (Statistics Canada, 2005a). CCHS cycle 3.1 employed the weighting methods of the Canadian Labour Force Survey (Statistics Canada, 1998) and considered initial weights, out-of-scope units, household and person non-response rates, and seasonal effects in providing a final set of weights.

3.4.1.3 Health Regions

At the time of the design of the CCHS cycle 3.1 sample there were 125 Health Regions across Canada. During data collection, some splitting and merging happened among health regions. This decreased the number of health regions to 122. The small population sizes of some health regions prompted a concern about the identification of individuals, thus a minimum value of 70,000 people was set for keeping or “collapsing” health regions together. After collapsing, the CCHS comprised 101 grouped geographic health regions across the country (Statistics Canada, 2005a). This research does not include the northern Territories (Yukon, Northwest Territories, and Nunavut) because of complex socioeconomic and cultural issues. Therefore, the final analyses involved 100 community health regions from the 10 provinces in Canada.

3.4.2 Access to CCHS cycle 3.1 data

A Public Use microdata file was accessible through the Social Science Data Centre at Queen’s University. Software developed by the Queen’s University Information
Technology Services and Documents Unit of Queen’s University Libraries, Queen’s WebInterface For SPSS (Data Analysis on the WWW Using QWIFS), allowed us to perform preliminary data analyses. These analyses included frequency and descriptive analysis, cross-tabulating independent variables against the Health Region variable, and collapsing and regrouping some categories of variables. To obtain ecological measures of variables, the desired variable was cross-tabulated against the Health Region variable (GEOEDPMF) and the percentage of people with the characteristics of interest was calculated within each health region.

All variables required for this research except the Rural-Urban status variable were obtained from this database. The Rural-Urban status variable was part of the CCHS Master File which was accessed through the Statistics Canada Research Data Centre (RDC) at Queen’s University. Access to the CCHS Master File required approval from Statistics Canada. Upon approval of the application, the candidate received permission to access CCHS cycle 3.1 Master file data and the Rural-Urban status variable was measured for each health region.

As a condition of using the RDC, researchers must comply with the Statistics Canada rules and regulations related to privacy and confidentiality. For example, to comply with Statistics Canada analysis privacy rules, raw data from the CCHS Master File were rounded to the nearest 50 and although some of the sample results showed that the estimate is zero, it did not necessarily mean that the true population value was zero.

### 3.5 Ascertainment of Outcome

The outcome of interest in this thesis is a measure of population health as indicated by the percentage of people with at least good self-rated health in each health region measured on a five point Likert Scale (Excellent, Very Good, Good, Fair, or Poor). Various studies have
used subjective indicators of health such as “self-rated health” as proxy measures of population health (Blakely et al., 2002; Subramanian et al., 2001) and this measure has proved to be an independent and accurate predictor of clinical health status (Idler & Benyamin, 1997). In this thesis, the data source for measurement of health is the General Health part of the CCHS 2005 cycle 3.1. All participants answered up to 10 questions about their general physical and mental health, including self-rated health. Variable self-rated health or as named in the (CCHS variable “GENEDHDI”) indicates the respondent’s health status based on his/her own judgment. Higher scores indicate positive self-reported health status (Statistics Canada, 2005a). The question in the final questionnaire of the CCHS cycle 3.1 is (Statistics Canada, 2005b):

<table>
<thead>
<tr>
<th>In general, would you say “YOUR” health is?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1- Excellent</td>
</tr>
<tr>
<td>2- Very good</td>
</tr>
<tr>
<td>3- Good</td>
</tr>
<tr>
<td>4- Fair</td>
</tr>
<tr>
<td>5- Poor.</td>
</tr>
<tr>
<td>Don’t Know, Refused</td>
</tr>
</tbody>
</table>

Figure 3.1: General Health Status Variable in the CCHS cycle 3.1

To estimate the general health status of each health region, the percentage of people who rated their health as good, very good and excellent was aggregated to create a good category and the percentage of people who rated their health fair or poor was aggregated to create a poor category. Results were treated as a continuous variable in subsequent analyses.
3.6 Ascertainment of Exposure

To measure income inequality across health regions, a range measure was used. Range measures compare two extreme categories and are reported as a ratio (e.g., the ratio of percentage of the poorest people over the richest). Lack of availability of individual level data was the main reason for choosing range measures instead of disproportionality measures such as the Gini coefficient in this thesis.

In the CCHS cycle 3.1 all respondents answered a question about their income. They were asked: What is your best estimate of YOUR total personal income, before taxes and other deductions, from all sources in the past 12 months? (Statistics Canada, 2005b)

During the data processing phase of the survey, responses were categorized into five groups: less than $15,000 per year; $15,000 to $29,999 per year; $30,000 to $49,999 per year; $50,000 to $79,999 per year; and more than $80,000 per year. Queen's WebInterface For SPSS (Data Analysis on the WWW Using QWIFS) software permitted calculation of the percentage of respondents in various income categories for each health region. The income inequality variable was calculated as a ratio of the percentage of people whose income was less than $15,000 to those who made more than $80,000 in each health region during the year before interview (i.e., 2004). This measure of income inequality was used as a continuous variable in subsequent analyses. The median value of this measure was 5 and this value was chosen as the threshold value for high inequality. If the ratio of percentage of people whose income was less than $15,000 to those who made more than $80,000 in each health region was more than 5, the region was labeled as high inequality, otherwise the region was labeled as low inequality (Table 3.2).
3.7 Covariates

In this ecological study, various covariates required consideration with respect to the relationship between the measure of income inequality (exposure) and the measure of population health (outcome).

All covariates were assessed as potential confounders and also effect modifiers. To perform subgroup analyses we categorized each covariate into different levels. The median value of each variable was used to establish a cut-off point for categorization (Table 3.2).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Median Values</th>
<th>Cut-off point Values</th>
<th>Sub-groups</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income Inequality</td>
<td>4.84</td>
<td>Ratio of 5</td>
<td>High Inequality</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Low Inequality</td>
<td>52</td>
</tr>
<tr>
<td>Absolute income</td>
<td>46.6</td>
<td>47%</td>
<td>Rich Regions</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Poor Regions</td>
<td>52</td>
</tr>
<tr>
<td>Health care utilization</td>
<td>87.6</td>
<td>88%</td>
<td>High Utilization</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>High Utilization</td>
<td>52</td>
</tr>
<tr>
<td>Sex</td>
<td>49.5</td>
<td>50%</td>
<td>Male Regions</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Female Regions</td>
<td>77</td>
</tr>
<tr>
<td>Age</td>
<td>15.6</td>
<td>16%</td>
<td>Senior Regions</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Non-senior Regions</td>
<td>61</td>
</tr>
<tr>
<td>Education</td>
<td>56.3</td>
<td>57%</td>
<td>High Education</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Low Education</td>
<td>54</td>
</tr>
<tr>
<td>Smoking</td>
<td>18.6</td>
<td>19%</td>
<td>High Smoking</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Low Smoking</td>
<td>56</td>
</tr>
<tr>
<td>Alcohol</td>
<td>21.6</td>
<td>22%</td>
<td>High Alcohol</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Low Alcohol</td>
<td>58</td>
</tr>
<tr>
<td>Rural/Urban</td>
<td>29.5</td>
<td>30%</td>
<td>Rural</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Urban</td>
<td>53</td>
</tr>
<tr>
<td>Social Capital</td>
<td>66.6</td>
<td>67%</td>
<td>High Social Capital</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Low Social Capital</td>
<td>54</td>
</tr>
<tr>
<td>Total Sample</td>
<td></td>
<td></td>
<td></td>
<td>100</td>
</tr>
</tbody>
</table>

Table 3.2: Subgroups of Variables
The covariates were divided into six different groups: 1) absolute income variables, 2) health care utilization variable, 3) Socio-demographic variables, 4) Substance use variables, 5) Rural/urban status variables, and 6) Social capital variables.

3.7.1 Absolute Income Variable

Clearly, absolute income is a covariate with a potential confounding effect as it is related to both outcome (self-perceived health) and exposure (income distribution). Indicators for absolute income include median household income, mean household income, and gross domestic product per capita. According to Statistics Canada (Statistics Canada, 2008, Table 202-0402), the median income of Canadian in 2006 was $26,900. Annual income of $30,000 was considered as the threshold of absolute individual wealth. To obtain an estimate of the absolute wealth of a health region, the percentage of relatively well-off people, whose annual income was more than $30,000, was calculated. The source of data was the INCEGPER variable in the CCHS 2005 cycle 3.1 (Table 3.1).

A continuous version of this variable was applied in these analyses. The variable had a median value of 47 per cent that was used as a cut-off point. For the sake of subgroup analyses and evaluation of the potential effect of this variable on the relationship of interest, health regions were divided into two groups, rich and poor regions. A region was labeled rich if more than 47 per cent of population made more than $30,000, otherwise the region was labeled as poor (Table 3.2).

3.7.2 Health Care Utilization

Health care utilization was also considered as a potential confounder (Shi et al., 2003). A variable labeled HCUE_1AA in CCHS cycle 3.1, asks the participants if they have a regular family doctor (Statistics Canada, 2005b). The percentage of people who have a regular medical doctor in each health region was calculated. The median value of this variable, 88
%) was used as the cut-off point for sub-grouping (Table 3.2). If more than 88 per cent of people had a regular family doctor, the region was labeled as high utilization; all other regions were considered low utilization (Table 3.2).

### 3.7.3 Socio-demographic Variables

Various socio-demographic factors can also affect the relationship of interest. Some existing studies have controlled for these variables: for instance, unemployment rates were controlled for in models developed by Wilkinson (1992), Shi et al. (2003), and Macinko et al. (2004); education (Muller, 2002); sex (Bobak et al., 2000; Pampel & Zimmer 1989); age (Bobak et al., 2000; Pampel, 2002), and percentages of visible minorities (Mclaughlin & Stokes, 2002).

Three socio-demographic factors of sex, age and education were considered in the analyses (Table 3.1). Variable EDUDR04 (Statistics Canada, 2005b) indicates the highest level of education acquired by the respondent. Education level of the respondents is classified into four categories: 1) less than secondary school graduation, 2) Secondary school graduation, 3) some post-secondary education, and 4) post-secondary degree/diploma. The percentage of people who have at least some post-secondary education (categories 3 and 4) was calculated for each health region, descriptive analyses showed that this variable had a median value of 57 per cent. The median values were used as a cut-off point for sub-group analyses. If more than 57 per cent of people had some post secondary education, that region was labeled as a “high education” region (Table 3.2).

### 3.7.4 Substance Use

Substance use including alcohol consumption (Osler et al., 2003) and smoking (Kennedy et al., 1996; Pampel, 2002) were considered as potential confounders. Lower shares of income may cause high levels of stress, which in turn leads to alcohol consumption and
smoking with subsequent negative health effects. Rates of smoking and alcohol consumption in each health region were calculated and considered in the analyses.

Variable SMKEDSTY (Statistics Canada, 2005b) in CCHS cycle 3.1 has been derived from a question in the final questionnaire:

<table>
<thead>
<tr>
<th>At the present time, do “YOU” smoke cigarettes daily, occasionally or not at all?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1- Daily</td>
</tr>
<tr>
<td>2- Occasionally</td>
</tr>
<tr>
<td>3- Not at all</td>
</tr>
<tr>
<td>Don’t Know, Refused</td>
</tr>
</tbody>
</table>

Figure 3.2: Smoking Variable in the CCHS cycle 3.1

To measure the population rate of smoking, the percentage of people who are daily smokers in each health region was calculated. The median value for this variable was 18.6 per cent. To create two levels of this variable, if more than 19 per cent of people smoked daily that region was labeled as “high smoking”, otherwise it was labeled as “low smoking” (Table 3.2).

Heavy drinking was assessed based on a question in the CCHS Cycle 3.1 which asks the participants whether they ever regularly drink more than 12 drinks a week (Statistics Canada, 2005b). The rate of heavy drinking was established by calculating the percentage of people who consumed more that 12 drinks regularly each week in each health region. If more than 22 per cent (based on the median value of 21.6 % for the variable) of people living in an area consumed more than12 drinks each week, that area was labeled “high alcohol” (Table 3.2).

3.7.5 Rural/Urban Status

Rural/Urban geographic status was defined using the GEOonDUR2 variable (Statistics Canada, 2005b). The rural population for 1981 to 2001 refers to persons living outside centres with a population of 1,000 and outside regions with 400 persons per square kilometer (Du
Residents of urban fringes were grouped in the urban groups and those in rural fringes were grouped in rural group (Statistics Canada, 2005a).

To assess the Urban/Rural status of each health region, the percentage of people who lived in rural regions in each region was calculated and a continuous variable was obtained. For the sake of subgroup analyses, this variable was categorized in two groups: 1) rural: if more than 30 per cent of population lives in rural regions, 2) urban: if less than 30 per cent of population live in rural regions. The median value of 30% for this variable was chosen as the cut-off point (Table 3.2). To address the second objective, the effect of this variable on the relationship of interest was evaluated in more detail.

### 3.7.6 Social Capital

Social capital or the level of civic engagement and interrelated trust (Last, 2007a), was measured by using the variable GENE_10. This variable is based on a question in the CCHS cycle3.1 (Figure 3.3).

<table>
<thead>
<tr>
<th>How would you describe your sense of belonging to your local community?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Would you say it is:</td>
</tr>
<tr>
<td>1- Very strong</td>
</tr>
<tr>
<td>2- Somewhat strong</td>
</tr>
<tr>
<td>3- Somewhat weak</td>
</tr>
<tr>
<td>4- Very weak</td>
</tr>
<tr>
<td>Don’t Know, Refused</td>
</tr>
</tbody>
</table>

Figure 3.3: Social Capital Variable in the CCHS cycle 3.1

The percentage of people who answered this question as “very strong” or “somewhat strong” was calculated for each health region. To create different levels of social capital for each health region and to perform subgroup analyses if more than 67 per cent (based on
median value of 66.6%) of population had a strong sense of belonging to their community, their health region was labeled as high social capital, otherwise they were classified as low social capital (Table 3.2).

3.8 Statistical Analyses

Most of the analyses in this thesis were conducted using Queen’s WebInterface For SPSS (Data Analysis on the WWW Using QWIFS), software developed by Queen’s University Information Technology services and Documents Unit of Queen’s University Libraries as an interface for using SPSS version 15.0. In addition, SAS version 9.1(SAS Institute, Cary, NC) was used for further data manipulations and analyses.

3.8.1 Descriptive Analyses

The first step in analysis was generating a description of the variables under study. For each variable considered, the range, extreme values, and measures of central tendency were estimated. The latter was used to establish cut-off points for subgroup analyses of the variables. To obtain an estimate of the relationship between covariates, a correlation matrix was generated.

3.8.2 Analysis of the Relationship between Exposure and Outcome of Interest

This part of analysis was conducted by plotting the exposure against the outcome in a visual analysis, examining the correlation between the exposure and outcome of interest statistically, and via a simple linear regression analysis to ascertain how much of the variance in the outcome was explained by the exposure after controlling for salient covariates. In order to control for the potential confounding effects of covariates on the relationship between income inequality and health, two analysis strategies were employed: 1) controlling for the
effect of covariates by partial correlation analyses, 2) constructing models with two independent variables.

3.8.3 Subgroup Analysis

Subgroup analyses were conducted to examine the relationship of interest within various subgroups, as well as to investigate the potential modifying effect of covariates. Two analysis strategies were employed for subgroup analyses: 1) exploring and comparing the relationship of interest within different levels of covariates (Table 3.2), 2) including an interaction term for each of the covariates with two main independent variables; income inequality and absolute income.

3.8.4 Collinearity Diagnosis

Collinearity refers to high correlation between independent variables (Last, 2000, p.35). A more specific term in regression analysis, multi-collinearity, is “a situation in which at least some of the independent variables are highly correlated with each other” (Last, 2000, p. 118). Multi-collinearity can produce inaccurate parameter estimates in the regression models and should be assessed before a final model is developed.

Two measures of collinearity are tolerance values and the variance inflation factor (Elliot, 2000; Freund & Little, 2000). Tolerance values for each variable in a regression analysis refer to the proportion of variance for each independent variable that is not shared with other variables. When tolerance values are 0, there is perfect collinearity between independent variables. Small tolerance values (<0.2) indicate possible collinearity (Elliot, 2000). The variance inflation factor addresses the same concept and in equivalent to 1/tolerance. High variance inflation factor values (>10) for a variable suggest the possibility of multi-collinearity (Freund & Little, 2000). Both tolerance and variance inflation factors for each
independent variable were measured by involving tol and vif options in the REG procedure of SAS software and the results were compared to cut-off points.

An alternative way to assess multi-collinearity is to examine the correlation matrices of the independent variables. They were assessed by incorporating the collin option in the REG procedure of SAS software. SAS Collinearity Diagnostic output produces condition indices of the correlation matrices and every condition index greater than 30 reflects moderate or severe collinearity. The number of condition indices in the output equals the number of variables. Every condition index contributes to some proportion of variance for each independent variable. Two or more variance proportions more than 0.5 suggest a collinearity problem (Freund & Little, 2000). In the analysis, the proportion of variables in every condition index greater than 30 was assessed.

3.8.5 Tests of Interaction

First-order interaction terms (two-factor products) were constructed for all independent variables with the measures of income inequality and absolute income. In linear regression analyses, these interaction terms were included in the models and their level of significance was assessed.

3.8.6 Outlier Detection, Influential Diagnosis

Observations that do not appear to fit a linear regression model, often referred to as outliers, can be quite troublesome since they can bias parameter estimates and make the resulting analyses less useful. For this reason, it is important to examine the results of a statistical analysis to ascertain if such observations exist (Freund & Little, 2000).

Statistical software packages provide different ways of identifying outliers. A simple way is to evaluate the size of the residual (predicted value of each observation minus its observed value), and studentized residuals (residuals divided by their standard errors) for each
observation. Residual and/or studentized residual values greater than 2 or less than -2 for an observation suggest a potential outlier.

R and influence options in the REG procedure of SAS software are also able to provide additional measures for outliers, their potential influence on the regression analysis and the indications for deletion of an observation. Among various measures of influential diagnosis in a SAS output, leverage indicates the extremeness of each observation and the impact of each observation in determining the model fit. If the leverage is larger than a set cut-off point, it indicates an important influence of the flagged observation on the observed estimate. Other helpful diagnostics, Cook’s D and DFFITS, are useful in deciding which observations should be deleted because of their substantial influence on the estimates (Belsley, Kuh, & Welsch, 1980, Freund & Little, 2000).

Definitions of different influential diagnoses and their corresponding cut-off points are presented in appendix A.

3.8.7 Model Building

The aim of model building was to find the best and most parsimonious regression model which could explain the maximum variations in the outcome (health status) with the least possible numbers of variables. To achieve this goal, various statistical methods exist. One is to identify important variables by their level of significance and to keep the most significant variables in the model. Another is to rank the models by descending values of R-square and to consider the models with fewer numbers of variables and higher R-square for the final model (Freund & Little, 2000). Statistical software packages are useful tools in model building based on these concepts. The REG procedure in SAS version 9.1 (SAS Institute, Cary, NC) was employed to determine which model is the “best” and most parsimonious
model. Since there was no indication of multi-collinearity, all independent variables were involved in model building.

3.8.8 Sensitivity analysis

To evaluate the robustness of the chosen range measure of income inequality a sensitivity analysis was performed. Three other range measures of income inequality focusing on different parts of the income distribution were calculated. These were extracted from INCEGPER variable (Statistics Canada, 2005b) in the CCHS 3.1 and were calculated as follows:

1) The ratio of the number of people whose annual income was less than $50,000 to those who made more than $50,000 in CHR.

2) The ratio of the number of people whose annual income was less than $30,000 to those whose income were more than $80,000 in each CHR.

3) The ratio of the number of people whose annual income was less than $30,000 to those whose income were more than $50,000 in each CHR.

These measures were substituted for the primary measure of income inequality and the newly acquired parameter estimates were compared to the primary inequality measure.
3.9 References


Data Analysis on the WWW Using QWIFS: Queen's WebInterface for SPSS, Social Science Data Centre, Queen's University. http://jeff- ab.queensu.ca/library/accessnew/


Chapter 4: RELATIONSHIPS BETWEEN INCOME INEQUALITY AND HEALTH: AN ECOLOGICAL CANADIAN STUDY.

This manuscript conforms to the specifications for submission to the peer-reviewed journal *Social Science & Medicine*

Authors:
Afshin Vafaei$^1$, Mark Rosenberg$^{1,2}$, William Pickett$^{1,3}$

1. Department of Community Health and Epidemiology, Queen’s University, Kingston, Canada
2. Department of Geography, Queen’s University, Kingston, Canada
3. Department of Emergency Medicine, Queen’s University, Kingston, Canada

Corresponding Author:
Afshin Vafaei
Department of Community Health and Epidemiology
Queen’s University, Kingston Ontario, Canada K7L3N6
Phone: +1 (613) 531-8702, +1 (416) 440-8923
Fax:
Email: 5av14@queensu.ca
Abstract

Background:
Despite many investigations, there is still debate on whether income inequality is a determinant of population health. Results of Canadian studies are conflicting and non-definitive; most indicate that there is no relationship between income inequality and health within Canada. There is a need for further investigation into the validity of the ‘relative income’ hypothesis in Canada.

Objectives:
The first objective of this research was to test the “relative income” hypothesis across Canadian health regions. This hypothesis implies that health status in populations is a function of income distribution. The second objective was to examine this hypothesis further in rural versus urban populations in Canada.

Methods:
This research involved ecological and cross-sectional analyses of the potential relationship between income inequality and self-rated health as a measure of population health. The source of the data was the Canadian Community Health Survey, CCHS 2005 cycle 3.1. The units of analysis were health regions defined by Statistics Canada. The percentage of people who rated their health status as good or excellent was considered as a proxy measure for population health status. The primary exposure (relative income inequality) variable was the ratio of people whose personal income was less than $15,000 to those who make more than $80,000 in the year preceding the survey (2004). The main covariates were ecological measures of socio-demographic variables, social capital, substance use behaviours (smoking and alcohol consumption), rural/urban status of the region, and absolute income in the region. Correlation analyses and simple and multiple linear regressions were performed to ascertain
the relationship between income inequality and population health after adjusting for important covariates.

**Results:**

The measure of income inequality alone appeared to explain 18 per cent of the variability in measures of health status in populations. However, after adding the measure of absolute income to the model, although 29 per cent of the variability was explained, the independent contribution of the inequality measure became non-significant. Linear regression models suggested that the absolute income variable alone was able to explain 30 per cent of the variance in health status in populations. Other variables with a statistically significant contribution to the final model were ecological measures of education and alcohol consumption.

The effect of rural/urban geographic status on the relationship of interest was similar to other covariates. This variable did not change the individual relationship between income inequality or absolute income and the measure of population health status. In both rural and urban regions, absolute income and education had positive effects on population health. In urban regions alcohol consumption was a significant negative contributor to population health status, whereas in rural regions, smoking status had a significant negative effect on population health status.

**Conclusion:**

Study findings indicated that higher aggregate absolute income was correlated with better population health. Regions with higher percentages of post-secondary educated people had better rates of health. Drinking behaviour was also correlated with lower population health status. The study findings have implications for public health, economic and social policies.

**Keyword:** Income inequality, Self-rated health, Canada, Population Health
**Introduction**

Various socioeconomic factors can determine health status in populations. Among them, income is one of the important social determinants of health (Sorlie et al., 1995). Higher income, particularly in developing countries often leads to healthier environments (Rodgers, 1979). In the early 1990s, Wilkinson described a new hypothesis to explain the relationship between relative income and health (the ‘relative income hypothesis’). In a series of published papers and books, he suggested that among developed countries the distribution of income in a geopolitical aggregate and not the absolute amount of wealth is a more reliable determinant of health (Wilkinson, 1992, 1994, 1999, 2005). The relative income hypothesis triggered a large amount of debate and numerous studies with conflicting interpretations and non-conclusive results. International studies and research on unequal societies such as the United States and the United Kingdom are mostly supportive of the hypothesis, whereas results of studies within countries with smaller income imbalances such as Canada (Laporte & Ferguson, 2003), Denmark (Osler et al., 2002, 2003), and Finland (Blomgren et al., 2004) are either negative or non-conclusive. It seems that the relative income hypothesis might only operate in large geographic regions (Wilkinson & Pickett, 2006).

In Canada the hypothesis has been tested within individual provinces (Phillips, 2002; Veenstra, 2002a, 2002b), but not in smaller regions across the country. The purpose of the current study was therefore to examine the ‘relative income’ hypothesis across all Canadian health regions. We conducted ecological research to test the potential relationship between a measure of income inequality and a measure of population health within Canada.
Methods

Study Design

To examine the relationship between income inequality and population health across Canadian health regions, an ecological analysis of the Canadian Community Health Survey (CCHS) cycle 3.1, 2005 data was conducted. Income distribution is only measurable at an aggregate level (Szklo & Nieto, 2007; Wilkinson, 1994); hence, an ecological design was the preferred choice for the current study. We measured all variables at the ecological level.

Data and Measures

Source of data

We used the Canadian Community Health Survey (CCHS) Version 3.1 conducted in 2005 (Statistics Canada, 2005). The CCHS Cycle 3.1 is a national level survey of about 130,000 persons conducted in 2004. The target population included household residents aged 12 and older in all Canadian provinces and territories. Exclusions were the populations on Indian Reserves or Crown Lands, those residing in institutions, full-time members of the Canadian Forces Bases, and those in some remote regions.

The survey had a multistage stratified cluster design in which the dwelling was the final sampling unit. The CCHS Cycle 3.1 employed rigorous strategies in weighting and obtaining a high response rate and achieved a household-level response rate of 84.9 per cent, a person-level response rate of 92.9 per cent, and a combined national response rate of 79 per cent. All variables required for this research except the Rural-Urban status variable were obtained from the Public Use Microdata File of the CCHS Cycle 3.1. This file was accessed through the Social Science Data Centre at Queen’s University, Kingston, Ontario. Rural/Urb
**Dependent variable**

The dependent variable in this research was a measure of health status of populations. A subjective indicator of health (self-rated health) was used to ascertain the health status of the population. In the CCHS Cycle 3.1 all participants rated their current status of their health as one of the following categories: excellent, very good, good, fair, and poor. To provide an estimate of general health status of each health region, the percentage of people who rated their current state of health as excellent, very good, and good was calculated and results were used in subsequent analyses as a continuous variable (one measure per health region).

**Independent variable**

A range measure (Harper & Lynch, 2006) of income inequality (the ratio of two extreme categories of income) was employed to ascertain the value of income inequality in each Canadian health region. The ratio of the percentage of people whose income was less than $15,000 to those who made more than $80,000 in each health region during the year before interview (i.e., 2004) was calculated in each Canadian health region and this measure of income inequality was used as a continuous variable in subsequent analyses.

**Covariates**

Various covariates required consideration with regard to potential confounding effects on the relationship of interest as they might be related to both outcome (perceived health) and exposure (income inequality). The most important covariate, absolute wealth of each aggregate, was estimated by calculating the percentage of people who had a personal income of more than $30,000 in the year preceding the survey (2004) in each Canadian health region. Variables considered a priori as potential confounders and controlled for in analyses were health care utilization, sex, age, education, smoking rate, alcohol consumption, rural/urban status, and social capital (Table 1).
Analytic Strategy

Most of the analyses in this thesis were conducted using Queen’s WebInterface For SPSS (QWIFS), software developed by Queen's University Information Technology services and Documents Unit of Queen’s University Libraries as an interface for using SPSS version 15.0 (Data Analysis on the WWW Using QWIFS). In addition, SAS version 9.1(SAS Institute, Cary, NC) was used for further data manipulation and analyses.

The range, extreme values and measures of central tendency of each ecological variable were estimated (Table 2). Median values of variables were used to establish a threshold to divide the observations into two categories. These categories were utilized in sub-group analyses of variables and tests of interaction. We used Pearson’s correlation to describe bivariate relationships between continuous variables. Simple and multiple linear regression analyses were used to examine relationships between the variables of interest while controlling for other factors, and to understand how much variation in Canadian population health status can be explained by the independent variable.

Ethics

The research used data from secondary sources already in the public domain. Before starting data collection and analyses ethics approval was obtained. To avoid any potential identification and to protect the privacy of study participants, the Statistics Canada rules and regulations related to privacy and confidentiality were followed at all times during the research. For example, to comply with Statistics Canada analysis privacy rules, raw data from the CCHS Master File were rounded to the nearest 50 and although some of the sample results showed an estimate of zero, it did not necessarily mean that the true population value was zero.
Results

Except for the relative income inequality variable, all other ecological variables were normally distributed. To make the relative income inequality variable suitable for linear regression analyses, we created a logarithmic transformation of this variable (Table 2).

Results of correlation analyses suggested that the dependent variable (health status in populations) had a statistically significant positive correlation with the absolute income and education variables. The correlation between the outcome variable and two ecological measures of sex and health care utilization were not significant. All other variables showed significant negative correlations with the outcome variable (Table 3). The correlation between the primary exposure variable (relative income inequality) and the outcome variable became non-significant after controlling for a measure of absolute income (Pearson correlation coefficients were changed from -0.43 (P<0.0001) to +0.09 (P=0.3816)). Simple linear regression analysis showed that the relative income inequality measure appears to explain 18 per cent of the variance in the population health, but after adding the measure of absolute income to the model, relative income inequality was not a significant predictor of variations in the population health. The measure of absolute income was able to explain 30 per cent of the variability of the measure health status in populations and its independent contribution stayed significant after adding the measure of relative income inequality to the model (Table 4).

Various strategies such as using the stepwise selection option and other properties in the REG Procedure in SAS version 9.1(SAS Institute, Cary, NC) were employed to identify important covariates and to build a final and the most parsimonious model (Freund & Little, 2000). Our final model showed that apart from absolute income, two variables, education and alcohol consumption measured at the ecological level have positive and negative
statistically significant effects on population health status, respectively. No second order interaction terms containing the variables in the final model (absolute income, education, and alcohol consumption) were significant. Based on tests of collinearity, multicollinearity between variables did not appear to be an issue in the regression analyses. Furthermore, we conducted outlier detection and influential diagnostics analyses. Health regions with extreme values did not have any significant influence on the results of regression analyses.

The rural/urban geographic status variable had an effect similar to other covariates on relationships of interest. When this variable was controlled for in the linear regression analyses, the individual contribution of income inequality and absolute income on population health status did not change (Table 4). In both rural and urban geographic regions the measure of relative income inequality became non-significant after adding absolute income to the model (Table 4, part C). In both rural and urban geographical regions, absolute income versus relative income, was a better predictor of health disparities in the population (in rural regions the adjusted $R^2 = 0.25$ vs. 0.09 and in urban regions 0.11 vs. 0.07; Table 5).

**Discussion**

Our national study systematically tested competing absolute income and relative income hypotheses in analyses of Canadian health regions. Previous Canadian research was mostly either non-conclusive (Sanmartin et al., 2003; Veenstra, 2002a) or non-supportive of the relative income hypothesis (Hou & Myles, 2005; McLeod et al., 2003; Veenstra, 2002b). The research demonstrated that compared to the relative income hypothesis, the absolute income hypothesis was better able to explain variations in the measure of population health status in Canada. In the regression analyses, when the measure of absolute income was considered as the main independent variable, the adjusted $R^2$ was higher compared to the models in which relative income was the main independent variable (0.30 in model 1b. versus 0.18 in model
The measure of absolute income remained a significant and independent contributor in the final model after controlling for potential confounders. This finding was similar to other studies on small geographic regions (Franzini et al., 2001), one thesis on Ontario public health units (Phillips, 2002), and research on countries with lower degrees of income inequality (Blomgren et al., 2004; Laporte & Ferguson, 2003; Wilkinson & Pickett, 2006).

The findings may be interpreted in two ways. Our units of analysis (Canadian health regions) were relatively small geographically and in terms of population size. In smaller regions, there is a lower likelihood of social heterogeneity, which is the main reason for the effect of income inequality on health (Wilkinson, 1997a). In smaller regions, population health is a function of absolute income not relative income (Wilkinson, 1997b). Second, there is also some debate on the nature of the relationship between income inequality and health. Speculating that the relationship between income inequality and health may depend on the level of inequality and be apparent only above a certain income inequality threshold (Asada, 2007). Canada, compared to some other western countries, namely the United States and the United Kingdom, has lower levels of income inequality (the Gini coefficient for Canada=0.32, for United States=0.45, and for the United Kingdom=0.34, (The world fact book, 2008)). The lower level of income inequality in Canada may be another reason for lack of support of the relative income hypothesis in our research.

Based on our final and optimal model (model 6a. in Table 4), the existence of a higher percentage of people with post-secondary education is associated with better population health status. It is not surprising that better educated people are also healthier, but we conclude that the more educated people in each region also affected the health of the whole community. Also, our model showed that the higher proportion of heavy drinkers in a region was a negative contributor to population health.
Rural/urban geographic status of a region was neither a confounder nor an effect modifier factor for the relationship between absolute income and population health in our study. Final models were constructed separately for each of rural and urban regions. In both regions, absolute income and education had positive effects on the measure of population health. In rural regions in contrast to urban regions, an ecological measure of smoking had a significant negative effect on population health while alcohol consumption had no effect.

The main methodological concern in this study was the choice of the relative income inequality measure. We did not have access to individual data required for calculating disproportionality measures such as the Gini coefficient and used a “range measure” based on an existing approach to assessment of income inequality (Harper & Lynch, 2006). Range measures are simple to calculate and interpret but do not take into account the weight of population size. To assess the sensitivity of our range measure, three other range measures which varied the ratio between the percentages of high and low income persons were also calculated. All correlation and regression analyses were repeated with these three new measures and no significant change in correlation coefficients and parameter estimates were observed. Our measure of income inequality appeared to be a reliable measure of relative income inequality.

Limitations of this research warrant comment. First, the cross-sectional study design did not permit a full exploration of temporality. Measures of absolute income and population health were measured at the same time and we could not establish if higher wealth in a region actually causes a higher percentage of people with good or excellent self-rated health. Second, since all variables were measured at an ecological level, we can not extend our conclusions to individuals. We showed that among Canadian health regions, more prosperous regions have higher percentages of people with good self-rated health, but one should not conclude that richer individuals inside a region are necessarily healthier. Finally, a lack of
available individual data did not permit the conduct of a multi-level analysis (CCHS public use data only provided the percentage of each category of self-rated health). Not performing a multi-level analysis increased the risk of missing the cross-level confounding effects between measure of income inequality at the area level and self-rated health at individual level. Our database (CCHS, 2005, cycle 3.1), though it employed a reliable sampling methodology, did not cover the entire Canadian population. Particularly, exclusion of on reserve Aboriginal People from the CCHS database, who are among the poorest and least healthy of the Canadian population, is a problem in generalizing the results of this research.

The study supported the absolute income hypothesis in Canada. Its findings have important implications for public health, economic, and social policies. Economic growth will improve population health in Canada; however, policies to improve population health that target low income individuals have the potential to ignore those whose health is poor, but their incomes are relatively high.

Longitudinal analyses are required to establish whether changes in absolute income will have an effect on the health status in populations. More research is needed to refute or approve the study findings. To establish a more comprehensive picture of relationships between income and health among Canadian populations, inclusion of on reserve Aboriginal People in future studies is warranted.

**Conclusion**

In summary, this study showed that absolute income compared to relative income inequality can better explain health disparities among the Canadian population. The study results showed that higher aggregate absolute wealth and higher percentages of people with post-secondary education are associated with better population health. Higher rates of alcohol consumption were also correlated with lower population health status. Policies on improving
income, education and lowering alcohol consumption will improve health. Since research on population health and the factors that determine better health within the Canadian population is still scarce, further studies are warranted that take into account the differences between relative income and absolute income status, Canada’s poorest and least healthy population groups, and longitudinal frameworks in contrast to cross-sectional frameworks.

**Acknowledgements**

Afshin Vafaei was funded by a Public Health and the Agricultural Rural Ecosystem (PHARE) Graduate Training Program scholarship, Canadian Institute of Health Research (CIHR).
References


Data Analysis on the WWW Using QWIFS: Queen's WebInterface for SPSS, Social Science Data Centre, Queen's University. http://jeff- ab.queensu.ca/library/accessnew/


<table>
<thead>
<tr>
<th>Name in CCHS</th>
<th>Label</th>
<th>Concept</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENEDHDI</td>
<td>Health</td>
<td>Population Health</td>
<td>Percentage of people with good, very good and excellent health in each Health Region.</td>
</tr>
<tr>
<td>INCEGPER</td>
<td>Inequality</td>
<td>Income Inequality</td>
<td>Ratio of number of people whose income is less than $15,000 to those who make more than $80,000 in each Health Region.</td>
</tr>
<tr>
<td>INCEGPER</td>
<td>Absolute</td>
<td>Absolute Income</td>
<td>Percentage of people whose income is more than $30,000 in each Health Region</td>
</tr>
<tr>
<td>HCUE_1AA</td>
<td>MD</td>
<td>Health Care Utilization</td>
<td>Percentage of people who have a regular medical doctor in each Health Region</td>
</tr>
<tr>
<td>DHHE_Sex</td>
<td>Sex</td>
<td>Sex</td>
<td>Percentage of males in each Health Region</td>
</tr>
<tr>
<td>DHHEGAGE</td>
<td>Age</td>
<td>Age</td>
<td>Percentage of people who are more than 65 years old in each Health Region</td>
</tr>
<tr>
<td>EDUEDR04</td>
<td>Education</td>
<td>Education</td>
<td>Percentage of people who have at least some post secondary education in each Health Region</td>
</tr>
<tr>
<td>SMKEDSTY</td>
<td>Smoking</td>
<td>Daily Smoking</td>
<td>Percentage of people who are daily smokers in each Health Region.</td>
</tr>
<tr>
<td>ALCE_6</td>
<td>Alcohol</td>
<td>Alcohol Consumption</td>
<td>Percentage of people who consume more that 12 drinks regularly each week in each Health Region.</td>
</tr>
<tr>
<td>GEOonDUR2</td>
<td>Rural</td>
<td>Rural /Urban Status</td>
<td>Percentage of people who live in rural regions in each Health Region.</td>
</tr>
<tr>
<td>GENE_10</td>
<td>Belonging</td>
<td>Social Capital</td>
<td>Percentage of people who rate their sense of belonging to their local community as strong in each Health Region.</td>
</tr>
</tbody>
</table>
Table 2. Mean, Minimum, Median, and Maximum Values of Variables Used in the Analysis

<table>
<thead>
<tr>
<th>By Health region</th>
<th>Mean</th>
<th>Median</th>
<th>Min.</th>
<th>Max.</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent reporting good health</td>
<td>87.6</td>
<td>88</td>
<td>80.3</td>
<td>92.5</td>
<td>12.2</td>
</tr>
<tr>
<td>Income Inequality (Ratio of less than $15,000 to more than $80,000)</td>
<td>7.15</td>
<td>4.84</td>
<td>1.47</td>
<td>38.0</td>
<td>36.5</td>
</tr>
<tr>
<td>Absolute Income (Percent with more than $30,000 earners)</td>
<td>45.3</td>
<td>46.6</td>
<td>28.3</td>
<td>58.6</td>
<td>30.3</td>
</tr>
<tr>
<td>Percent with a regular Family Doctor</td>
<td>86.9</td>
<td>87.6</td>
<td>67.5</td>
<td>97.2</td>
<td>29.7</td>
</tr>
<tr>
<td>Percent of males</td>
<td>49.5</td>
<td>49.5</td>
<td>47.5</td>
<td>52.4</td>
<td>4.90</td>
</tr>
<tr>
<td>Percent older than 65 years</td>
<td>15.5</td>
<td>15.6</td>
<td>7.40</td>
<td>23.7</td>
<td>16.3</td>
</tr>
<tr>
<td>Percent with at least secondary education</td>
<td>56.4</td>
<td>56.3</td>
<td>42.4</td>
<td>71.7</td>
<td>29.3</td>
</tr>
<tr>
<td>Percent of daily smokers</td>
<td>18.3</td>
<td>18.6</td>
<td>7.30</td>
<td>27.0</td>
<td>19.7</td>
</tr>
<tr>
<td>Percent of heavy drinkers</td>
<td>21.4</td>
<td>21.6</td>
<td>5.60</td>
<td>43.2</td>
<td>37.6</td>
</tr>
<tr>
<td>Percent of rural population</td>
<td>29.9</td>
<td>29.5</td>
<td>0.00</td>
<td>82.0</td>
<td>82.0</td>
</tr>
<tr>
<td>Percent with strong sense of belonging</td>
<td>65.3</td>
<td>66.6</td>
<td>23.1</td>
<td>95.2</td>
<td>72.1</td>
</tr>
</tbody>
</table>
Table 3. Pearson Correlation Matrix of Variables

<table>
<thead>
<tr>
<th></th>
<th>Health</th>
<th>Inequality</th>
<th>Absolute</th>
<th>Physician</th>
<th>Sex</th>
<th>Age</th>
<th>Education</th>
<th>Smoking</th>
<th>Alcohol</th>
<th>Rural</th>
<th>Belonging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health</td>
<td>1.00</td>
<td>-0.43§</td>
<td>0.55§</td>
<td>-0.06</td>
<td>0.03</td>
<td>-0.38§</td>
<td>0.50§</td>
<td>-0.39§</td>
<td>-0.32‡</td>
<td>-0.44§</td>
<td>-0.26†</td>
</tr>
<tr>
<td>Inequality</td>
<td>1.00</td>
<td>-0.77§</td>
<td>0.06</td>
<td>-0.20*</td>
<td>0.32†</td>
<td>-0.37‡</td>
<td>0.17</td>
<td>0.06</td>
<td>0.58§</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>Absolute</td>
<td>1.00</td>
<td>0.01</td>
<td>0.16</td>
<td>-0.50§</td>
<td>0.51</td>
<td>-0.25*</td>
<td>-0.20</td>
<td>-0.63§</td>
<td>-0.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physicians</td>
<td>1.00</td>
<td>-0.37‡</td>
<td>0.15</td>
<td>0.04</td>
<td>-0.26‡</td>
<td>-0.02</td>
<td>-0.003</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>1.00</td>
<td>-0.38§</td>
<td>-0.34‡</td>
<td>0.30†</td>
<td>0.03</td>
<td>0.18</td>
<td>-0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.00</td>
<td>0.32†</td>
<td>0.04</td>
<td>0.24*</td>
<td>0.37§</td>
<td>0.26†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>1.00</td>
<td>-0.60§</td>
<td>-0.19</td>
<td>-0.65‡</td>
<td>0.30†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td>1.00</td>
<td>0.29†</td>
<td>0.38§</td>
<td>-0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol</td>
<td>1.00</td>
<td>0.30†</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>1.00</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p<0.05, † p<0.01, ‡ p<0.001, § p<0.0001
<table>
<thead>
<tr>
<th>Models</th>
<th>Independent variable(s)</th>
<th>β</th>
<th>Stand. β</th>
<th>Adj. R²</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 a. Inequality</td>
<td></td>
<td>-1.50†</td>
<td>-0.43</td>
<td>0.18</td>
<td>-4.76</td>
<td>0.0001</td>
</tr>
<tr>
<td>1 b. Absolute Income</td>
<td></td>
<td>0.18†</td>
<td>0.55</td>
<td>0.30</td>
<td>6.25</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>2. Absolute + Inequality</td>
<td>(Inequality)</td>
<td>0.50*</td>
<td>0.14</td>
<td>0.88</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>(Absolute)</td>
<td>(Absolute)</td>
<td>0.22†</td>
<td>0.67</td>
<td>4.11</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>3a. Inequality + Covariates</td>
<td>-0.67*</td>
<td>-0.20</td>
<td>0.34</td>
<td>-1.63</td>
<td>0.1100</td>
<td></td>
</tr>
<tr>
<td>3b. Absolute + Covariates</td>
<td>0.12†</td>
<td>0.37</td>
<td>0.40</td>
<td>3.18</td>
<td>0.0100</td>
<td></td>
</tr>
<tr>
<td>4. Inequality + Absolute + Covariates</td>
<td>(Inequality)</td>
<td>0.47†</td>
<td>0.14</td>
<td>0.84</td>
<td>0.4100</td>
<td></td>
</tr>
<tr>
<td>(Absolute)</td>
<td>(Absolute)</td>
<td>0.15†</td>
<td>0.47</td>
<td>2.81</td>
<td>0.0100</td>
<td></td>
</tr>
<tr>
<td>5a. Inequality + Rural</td>
<td>-0.89*</td>
<td>-0.26</td>
<td>0.22</td>
<td>-2.23</td>
<td>0.0241</td>
<td></td>
</tr>
<tr>
<td>5b. Absolute + Rural</td>
<td>0.15†</td>
<td>0.45</td>
<td>0.31</td>
<td>4.16</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>6a. Absolute + Alcohol + Education</td>
<td></td>
<td>0.12†</td>
<td>0.37</td>
<td>0.39</td>
<td>3.99</td>
<td>0.0001</td>
</tr>
<tr>
<td>6b. Absolute + Alcohol + Smoking + Belonging</td>
<td></td>
<td>0.13†</td>
<td>0.40</td>
<td>0.39</td>
<td>4.73</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>6c. Absolute + Alcohol + Smoking + Education</td>
<td></td>
<td>0.13†</td>
<td>0.38</td>
<td>0.39</td>
<td>4.10</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

* β, t, and p for Income Inequality variable.
† β, t, and p for Absolute Income variable.
Table 5. Results of linear regression analysis of association between income inequality, absolute income and health within different levels of rural/urban status (Cut off=30)

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>β</th>
<th>Standardized β</th>
<th>Adjusted R²</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
</table>
| A) Independent variable= Income Inequality
| Rural regions  | 47  | -1.14 | -0.34          | 0.09        | -2.39 | 0.0209 |
| Urban regions  | 53  | -1.12 | -0.29          | 0.07        | -2.19 | 0.0332 |

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>β</th>
<th>Standardized β</th>
<th>Adjusted R²</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
</table>
| B) Independent variable= Absolute Income
| Rural regions  | 47  | 0.18 | 0.52           | 0.25        | 4.05 | 0.0002 |
| Urban regions  | 53  | 0.14 | 0.36           | 0.11        | 2.76 | 0.0080 |

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>β</th>
<th>Standardized β</th>
<th>Adjusted R²</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
</table>
| C) Independent variables= Income Inequality+ Absolute Income
| Rural regions  | 47  | 0.57 | 0.17           | 0.25        | 0.83 | 0.4119 |
| Urban regions  | 53  | 0.22 | 0.06           | 0.10        | 0.23 | 0.8213 |

* β, t, and p for Income Inequality variable.  
† β, t, and p for Absolute Income variable.
Chapter 5:  
General Discussion

The purpose of this thesis was to test the relative income hypothesis of population health versus the absolute income hypothesis in the Canadian population. The research represents the first such study to involve all Canadian health regions in testing the relationship of interest. This was a main strength of the thesis. The results showed that the absolute income hypothesis can better explain health disparities among Canadian health regions.

The choice of units of analysis might be one reason for the better performance of the absolute but not the relative income hypothesis in this study. Previous studies of relationships between income inequality and population health mostly compare large geographical regions (countries, provinces, states) and were more supportive of the relative income hypothesis. Our units of analysis (Canadian health regions) were relatively small geographically and in terms of population size. Wilkinson (2000) argues that in moving from larger to smaller regions, absolute income becomes a more important predictor, and relative income a weaker predictor, of mortality. The same observation could be extended to measures of morbidity and health status. The research in this thesis appears to support this argument that examines the relationship between income and population health status.

The main limitations of the research were associated with the use of a secondary database. We were also unable to establish temporality due to the cross-sectional nature of data, and the problems of generalizing the ecological findings to individuals within the ecological units. Strengths of the thesis were including all Canadian health regions in the study and also our test of the effect of rural/urban status on the relationship of interest.
The discussion is organized as follows. First, the findings will be presented and compared with the existing literature. Next, limitations and strengths of this study will be addressed. Potential contributions of the study and its implications for public health policies will be discussed. Finally directions for future research, recommendations and conclusion are provided.

5.1 Main Findings

The research showed that after controlling for absolute income, the effect of relative income inequality on population health is eliminated. This finding is the main and most important finding of this thesis, and was established with both correlation and regression analyses. Income inequality, measured by the ratio of percentage of people whose income was less than $15,000 to those who made more than $80,000 annually in each Canadian health region, appears to explain 18 per cent of the variability in the population’s health. However, when a measure of absolute income, the percentage of people whose income was more than $30,000 in each Canadian health region, was included in the linear regression analyses, although 29 per cent of the variability was explained, the independent contribution of the relative income inequality measure became non-significant. This finding was consistent with other studies on smaller geographic regions (Massing et al., 2004; Shi et al., 2005; Stanistreet et al., 1999) and an unpublished Queen’s University thesis (Phillips, 2002).

A second important finding was evidence in support of the absolute income hypothesis within Canada. Results of the regression analyses revealed that when absolute income was the main independent variable, the adjusted R² was higher compared to the models in which relative income was considered as the main independent variable (0.30 vs. 0.18, model 1b. versus model 1.a, table B.9). Among Canadian health regions and based on analyses using CCHS 2005, cycle 3.1 data, the absolute income variable alone explained 30 per cent of the
variance in our measure of population’s health, independent of other covariates. This finding closely matches results from another thesis conducted by Phillips (2002) at Queen’s University. In that study, conducted on Ontario public health units, 25.1 per cent of the variance in mortality was explained by median income. Construction of additional models showed that “absolute income” remained a significant predictor of population health after controlling for the effects of other covariates individually (Table B.8) and together (models 3b, 4, and 5b in table B.9).

Our final model showed that apart from absolute income, among all covariates involved in the model two variables; education and alcohol consumption measure at the ecological level; have positive and negative effects on population health, respectively (Figure, B.3).

Various measures of income inequality may reflect different distribution patterns and yield different results (Bobak et al., 2000; Gold et al., 2001). To assess the sensitivity of the primary measure of income inequality, three other measures of income inequality were employed in sensitivity analyses (section 3.8.8). Each showed highly significant positive correlations with each other (Pearson $\rho=0.90$ to 0.99, $P=<0.0001$) and linear regression results indicated that employing these different measures did not change the parameter estimates. These findings are consistent with at least one other study (Kawachi & Kennedy, 1997) that found the choice of measure of income inequality does not affect the relationship with mortality, and measures are typically highly correlated with each other. However, another study (Kennedy et al., 1996) in contrast to the present study, found that the relationship between income inequality and health was not the same among two different inequality measures.

To address the second objective of the thesis, the effect of rural/urban status was evaluated with regard to the relationship of interest, and it proved to be similar to other covariates. After controlling for this variable, the individual contribution of income inequality (Table B.5) and
absolute income (Table B.7) on health status in populations did not change. When the regions were divided into two geographic groups with regard to their rural/urban status, in both regions the measure of income inequality became non-significant after adding absolute income to the model (Table B.10, part C). In both regions, absolute income, versus relative income, was a better predictor of health disparities in the population (in rural regions the adjusted $R^2=0.25$ vs. 0.09 and in urban regions 0.11 vs. 0.07; Table B.10).

The final models were slightly different in rural versus urban regions. In both regions, absolute income and education showed positive effects on the measure of population health. In urban regions, alcohol consumption had a significant negative effect on population health while in rural regions the only covariate with a significant negative effect on population health was smoking (Figure B.4).

5.2 Limitations of the Thesis

5.2.1 Design Limitations

Four limitations of this research are related to its application of cross-sectional data, the ecological design, and high observed correlations between income inequality and absolute income measures. Similar to other cross-sectional studies, this study was not capable of providing an assessment of temporality. The study showed a correlation between absolute income and population health across Canadian health regions, but was not able to establish either causality or rule out spurious correlations.

Units of analysis in this research were Canadian health regions and all variables were measured at the aggregate level. Conclusions should be drawn only at the aggregate level. The study showed that the more prosperous Canadian health regions are healthier but it did not necessarily mean that richer individuals were also healthier individuals irrespective of the region in which they live. Another limitation of our ecological approach is the possibility of
cross-level confounding between a contextual exposure (measure of income inequality in an area) and individual level exposures and outcomes (Blakely et al., 2002). Use of multi-level analyses is a possible solution for this limitation, but unavailability of individual data prevented us from applying this analytical approach.

Tests of collinearity showed very low probability of collinearity between independent variables (section B.6). However, the income inequality measure showed a highly significant negative correlation with the measure of absolute income ($\rho=-0.77, p<0.0001$), Table B.2). This high correlation should be taken into account in interpretation of findings of this research.

5.2.2 Limitations of the Dataset

The study was constrained by limitations of the CCHS secondary dataset. The CCHS dataset did not cover all of the Canadian population in its sample frame. Some aboriginal communities, military personnel, penitentiary populations and some very remote regions were not included in the CCHS sample. Lack of inclusion of reserve aboriginal people in the data was particularly problematic in this study. Aboriginal people are among the poorest and the least healthy of the Canadian population and values for both their income and health fall in the lowest tails of both distributions (MacMillan et al., 1996; Waldram et al., 2006). Inclusion of aboriginal people in the study of the relative income versus absolute income in Canada would provide more definitive results.

5.2.3 Policy Limitations

Policy limitations should be recognized. One rationale behind this study was to provide guidance for health and economic policies. In each region, certain proportions of the population will rate their health as poor, regardless of any health policy based on either absolute or relative income. This means that trying to affect population health by only
targeting regions where the high percentage of the population is in poor health through income policies might have less effect if the population is relatively well-off financially. While policies to improve population health that target low income individuals have the potential to ignore those whose health is poor, but their incomes are relatively high. Speculatively in a free market country such as Canada, the political parties are limited as to how much they can affect the income distribution of the population without losing voter confidence.

5.3 Methodological Considerations

The first issue to be considered methodologically is the measure of relative income inequality employed. There is no gold standard measure of income inequality. I chose to use a “range measure” that was simple to calculate and interpret, but does not adjust for the population size (section 2.3.3). Calculation of disproportionality measures such as the Gini coefficient requires individual data and this allows one to consider and weight the population size. The main reason for employing a range measure in this study was the lack of availability of individual data. To the best of my knowledge, no study exists that proves which measure is a more valid indicator of income inequality, but it is possible that the application of other measures with the same data might give rise to different results.

Selection bias (error due to systematic differences in characteristics between those who take part in a study and those who do not (Last, 2001, p. 166)) can be a problem in this study as in any other survey. However, the CCHS cycle 3.1 employs a very sophisticated sampling and data collection design (section 3.4.1) and shows a high response rate of 79 per cent which minimizes the possibility of selection bias with the exception of the excluded sub-populations.
Potential confounders in this study were chosen based on the literature (section 3.7) and the obtained correlation matrix (Table 4.2). The investigator considered most important covariates in the regression analyses; however, variables such as ethnicity and immigration status with potential confounding effects on the relationship of interest were not included in the analyses because of the way they were reported in the CCHS cycle 3.1.

5.4 Strengths of the Thesis

Some Canadian studies have examined the relative income hypothesis on small geographic regions inside a province. Studies within Canada on small geographic regions are scarce. Except for two studies by Veenstra on coastal communities of British Columbia (2002b) and health districts of Saskatchewan (Veenstra, 2002a) and two theses on this topic using Ontario public health units (Phillips, 2002; Xi, 2003), the present study was, to the best of my knowledge, the first study that analyzed relationships between health and relative income across all Canadian health regions. Analyzing the effect of the rural/urban gradient on the relationship of interest was also a strength of the present study even though the rural/urban gradient did not turn out to be statistically significant in any of the models. The data source of this study, the CCHS 2005, cycle 3.1 was a reliable source with a large sample size and very robust sampling methodology (Statistic Canada, 2005).

Other strengths of this thesis are related to the analytic methodology. The research systematically analyzed both the relative and absolute income hypotheses among Canadian health regions and took into account the potential confounding effects of various ecological variables (at an ecological level) in regression analyses (Tables B.5, B.8, B.9).

5.5 Contribution and Policy Implications

This research provides important evidence that is new for Canada. The first knowledge gap was the scarcity of Canadian studies. There are numerous international studies which
involve Canada and a few mixed United States and Canadian studies, but pure Canadian studies have not been conducted. Existing Canadian studies compared metropolitan regions, provinces, or regions within a particular province. This thesis was the first study that tested the relative income hypothesis across all Canadian health regions. Another knowledge gap relates to the results of the existing Canadian studies. The results of existing Canadian studies are conflicting and inconclusive with an inclination to support the absolute income hypothesis. This thesis provides additional evidence in support of the validity of the absolute income hypothesis in Canada.

This research has important implications for economic, health, and social policies. The results suggest that an increase in income could improve one important indicator of population health. Any health promotion policy should also consider the impact of economic growth on health. The next important variable contributing to improving health based on the final model (Table 4) was education. Communities with higher proportions of highly educated people showed a better rate of population health. Drafting a social policy to provide better educational opportunities is likely to have a positive impact on health status of populations (Raphael, 2003). This should come as no surprise given the generally strong positive correlations between income and education levels both at the individual level and at aggregate scales of analysis. The last important covariate in the final model was alcohol consumption. Higher percentages of heavy drinkers (those who regularly drink more than 12 drinks each week) by region will have ill effects on population health. Public health measures to educate people to reduce their alcohol consumption may improve population health (Edwards et al., 1994).
5.6 Future Research Directions

There is a large and growing body of literature on the relationship between income and health; however, few studies have investigated the relationship in one single country and even fewer studies have been conducted in Canada. Further research on this topic in Canada, especially on small geographical regions such as health regions or counties, is warranted.

This cross-sectional, ecological research was not able to explore temporality. Statistics Canada provides ongoing data on health-related subjects in the form of time-series. To assess temporality, further research might use the National Population Health Survey (NPHS) longitudinal file and see if changes in relative income or absolute income would consequently lead to changes in population health. The other way to assess temporality of the relationship between income and health is to extract health data (outcome) from recent data and income data (exposure) from previous years. By assessing temporality, research would be one step closer whether income does have an impact on health.

One limitation to this research was the CCHS exclusion of on reserve aboriginal people. Surveys on aboriginal health (Aboriginal Peoples Survey) exist and can be used in future research on that subset of Canadian population (Statistics Canada, 1995).

There are other methodological options to address the question of the relationship between income and health. Other measures of population health such as mortality rates and different measures of income inequality may be used. Some argue that relationships between income inequality and health are intrinsically multilevel (Subramanian et al., 2001). In this research all variables were at one level (an ecological level), therefore single level regression analyses were performed. Further research might employ a multilevel approach by using both aggregate and individual data.
5.7 Recommendations and Conclusion

Income is one of the most important social determinants of health. There is still a debate on whether absolute income or relative income has more effect on the population health of industrial countries. Each hypothesis demands particular social and health policies. This research concluded that among health regions of Canada, absolute income compared to relative income may better explain variations in self-perceived health of Canadian populations. Further research with different methodologies and time series data applied on the same population are required to confirm or reject the findings of this thesis.
5.8 References


### Appendix A: Influential and Deletion Diagnostics

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
<th>Cut-off point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residual</td>
<td>Predicted value of each observation minus its observed value ($e_i$).</td>
<td></td>
</tr>
<tr>
<td>Studentized Residual</td>
<td>Residuals divided by their standard errors ($e_i / S$).</td>
<td>&lt;=-2 or 2</td>
</tr>
<tr>
<td>Leverage</td>
<td>Measures the extremeness of the $i$th observation in the range of $X$ values.</td>
<td>&gt;2($k+1$)/n</td>
</tr>
<tr>
<td>Cook’s D</td>
<td>Measures the extent to which the regression coefficient change when a particular observation in question is deleted.</td>
<td>&gt;1</td>
</tr>
<tr>
<td>DFFIT</td>
<td>Change in the predicted value for the $i$th observation without using the $i$th observation.</td>
<td>&gt;2*$\sqrt{(P/n)}$</td>
</tr>
<tr>
<td>Bfbeta</td>
<td>Change in beta value after deleting the observation.</td>
<td>&gt;2/$\sqrt{n}$</td>
</tr>
</tbody>
</table>

K = number of independent variables  
$n$ = number of observations  
$P$ = number of all variables
Appendix B: Full Study Results

B.1 Descriptive Statistics and Correlation Analysis

All variables in this thesis were measured at the ecological level and provided a summary measure for individual health units. Table B.1 summarizes the descriptive statistics of these variables. Preliminary descriptive analyses revealed that all variables except the income inequality variable were normally distributed across health regions (Figures B5, B6). To make the income inequality variable suitable for linear regression analyses, a logarithmic transformation of this variable was employed (Figures B7, B8).

For most variables the mean and median measures were very close to each other (Table B.1). It suggests that choosing the median over the mean for the cut-off values in subgroup analyses likely has little influence on the results. The dependent variable (health status) had a statistical significant positive correlation with absolute income and education variables. All other variables except sex and health care utilization (correlations between them and the dependent variable were non-significant) showed significant negative correlations with the outcome variable (Table B.2).

B.2 The Relationship between the Primary Exposure and Outcome of Interest

The exposure (income inequality) and outcome (health) variables were significantly correlated with each other (Pearson $\rho =-0.43$, $p<0.0001$), (Table B.2), and the relationship was linear as is illustrated by figure 5. There was considerable scatter around the summary function. The measure of inequality alone was able to explain 18 per cent of total variance in the outcome (Table B.3).
Correlation analyses were repeated by controlling for individual effects of each variable. The exposure and outcome variables remained significantly correlated except when the correlation was controlled for by a measure of absolute income. By controlling for absolute income the Pearson correlation coefficients changed from \(-0.43\) (\(P<0.0001\)) to \(+0.09\) (\(P=0.3816\)) (Table B.4).

After adding the covariates individually to the main model in which the primary independent variable was relative income inequality variable and dependent variable was a measure of population health, relationships stayed significant except when the measure of absolute income was entered in the model. After controlling for the effect of absolute income on the relationship between income inequality and health, the latter relationship was not significant anymore (Table B.5).

**B.3 The Exposure-Outcome Relationship within Different Levels of Covariates:**

Results of these subgroup analyses showed that the betas for income inequality and their corresponding p-values varied considerably within different levels of absolute income, health
care utilization, sex, age and education. This suggests the possibility of effect modification at the ecological level (Table B.3). However, after performing tests of interaction between the income inequality variable and each of the individual covariates, the only significant result was for the interaction between sex and inequality (p=0.03). The beta value for female regions was negative (Table B.3), whereas the betas for male regions were positive, a finding that is also shown graphically in figures B9 and B10. The relationship of interest did vary for the two levels of sex.

B.4 The Alternative Hypothesis, Absolute Income Hypothesis:

To examine whether across Canadian health regions, the absolute income hypothesis better explains variations in the chosen indicator of health status of populations, a second set of models was constructed in which absolute income was employed as the primary exposure variable. Correlation analyses showed high and significant positive correlation between health status and absolute income (Pearson ρ=0.55, p<0.0001), (Table B.2). The plot of absolute income against health status demonstrated a positive linear relationship (Figure B.2).

Figure B.2: Scatter plot of Population Health against Absolute Income.
The results of linear regression analyses (Table B.6) indicated a higher adjusted $R^2$ compared to the relative income measure (0.30 vs. 0.18) and the parameter estimate for absolute income was also highly significant ($p<0.0001$). Measures of absolute income and health status in populations remained significantly correlated after each covariate was individually controlled for in correlation analyses (Table B.7). After adding covariates individually to models in which the measure of absolute income was kept as the primary independent variable, no major changes in the betas for absolute income were observed (Table B.8). All models showed a highly statistically significant effect ($P<0.0001$) indicating that the relationship between absolute income and population health did not change after controlling for other existing ecological covariates.

**B.5 The Absolute Income Hypothesis within Different Levels of Covariates:**

Results of regression analyses summarizing the relationship between absolute income and population health with salient subgroups are presented in Table B.6. Parameter estimates for two variables (sex, age, and education) varied within subgroups. There was a possibility of effect modification. Tests of interaction between all variables were performed. Only two variables (sex and age) demonstrated significant interactive effects with absolute income, and the $p$-values for both of these interaction terms were 0.03. Scatter plots showing the relationship between absolute income and population health for different levels of sex and age are shown in figures B.11 through B.14. The plots for the sex strata show different slopes, whereas within the age strata, the slopes are very similar.

**B.6 Tests of Collinearity**

Results of tests of collinearity indicated that tolerance values for all independent variables were larger than 0.2 and all variance inflation factor values were smaller than 10. This
suggests a very low probability of collinearity. In collinearity diagnostics output, three condition indices had values greater than 30. Only one of them contributed to more than 0.5 proportions of variations of only one independent variable (absolute income). Based on these diagnostics, multicollinearity did not appear to be a concern.

**B.7 Regression Diagnostics**

A first step in outlier detection was establishing cut-off points for both outlier and influential diagnostics. Using a standard formula for cut-off points (Freund & Little, 2000). (Appendix A) and considering that the research data contained 100 observations, 10 independent variables and one dependent variable, the following values were obtained:

<table>
<thead>
<tr>
<th>Diagnostics</th>
<th>Calculation</th>
<th>Cut-off point</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leverage (h)</td>
<td>$2\times(10+1)/100$</td>
<td>0.22</td>
</tr>
<tr>
<td>DFFIT</td>
<td>$2\times\sqrt{(11/100)} = 2\times0.33$</td>
<td>0.66</td>
</tr>
<tr>
<td>BFBETAS</td>
<td>$2/\sqrt{100}$</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Observations for one health region (Chatham-Kent) were suggestive of being an influential outlier. This observation had a residual value of 4.05, a studentized residual of 2.36 (>2), and a DFFITS value of 0.81 (slightly larger than the cut-off point of 0.66). Values of other diagnostics, jackknife and BFBETAS, were also high but leverage and Cook’s d values were smaller than the standard cut-off points for this data set (0.1, and 0.061 respectively).

The Chatham-Kent observation and seven other health regions (zone 3 of Nova Scotia, Lac St. Jeans, Gaspésie-Îles-de-la-Madeleine, Palliser Health Region, Kootenay Bound, North Vancouver Islands, and North-Western Health Region of Ontario) had at least one high residual or influential diagnostics indicator. These were manually investigated for the possibility of data entry errors, and none was found. To control for potential influential effects of these observations, the final model was re-tested after deleting these variables. No
major change in parameter estimates and adjusted $R^2$ values was noted. As expected, the greatest change in beta for absolute income (from 0.12 to 0.11) happened after deleting the Chatham-Kent observation and $R^2$ increased from 0.390 to 0.4120. Since the change was evaluated as minimal, this observation was kept in the analysis.

**B.8 Building of a Final Model:**

The first step in model building was the identification of significant covariates. Model 2 (Table 4.9) shows that apart from the income variables, the covariates of education, alcohol and sex are important variables in explaining our measure of general population health. Model 4 showed that after involving all variables, the income inequality measure was no longer significant ($p=0.41$). Meanwhile the absolute income variable stayed significant ($p=0.01$). Subsequent models were established by using the R-square selection method as available in SAS procedures. In models labeled under number 6 (Table B.9), the adjusted $R^2$ values were high and very close, all involved the absolute income variable (further proof for the validity of the absolute income hypothesis versus the relative income hypothesis in these data), and all models were significant ($p$-values for F statistics in all models were all <0.0001).

The optimal model is 6a. (Figure B.3) because:

1) it is the most parsimonious, in that it has only three variables compared to four variables in models 6b., and 6c., and five variables in models 6d., and 6e. 2) in contrast to others, this model does not include non-significant covariates such as “smoking” and “belonging”. 3) the model has positive parameter estimates for absolute income and education and negative estimates for alcohol, which shows the first two variables will improve population health, whereas alcohol has ill effects on health. This finding is consistent with other studies (Massing et al., 2004; Phillips, 2002; Shi et al., 2005; Stanistreet et al., 1999). 4) the model
shows a high value of $R^2$ and can explain about 40 per cent of the variability in the outcome.  

5) no first-order interaction terms between the variables involved in the model were significant.

\[
\text{Health Status} = 78.2 + 0.12 \text{ absolute} + (-0.08 \text{ alcohol}) + 0.1 \text{ education}
\]

Or alternatively, by using standardized coefficients as:

\[
\text{Health Status} = 0.37 \text{ absolute} + (-0.2 \text{ alcohol}) + 0.28 \text{ education}
\]

The model explains about 39% of the variability in the outcome.

According to this model (first version) at the ecological level, if the percentage of people who make more that $30,000 each year increases by 1 unit (here 1%), there would be 0.12 addition to the percentage of people with at least good self-rated health. Also, if the percentage of post-secondary educated people increases by 1 unit, we may predict 0.1 increase to the percentage of people with at least good self-rated health in each health region. For each one unit (here 1%) increase in the percentage of alcohol consumers in each health region, there would be 0.08 decrease in the percentage reporting this measure of population health as at least good. Since this model does not involve any significant interaction terms, there is no need to report the results within different levels of the variables.

**B.9 The “relative income” and “absolute income” Hypotheses in Rural versus Urban Populations**

To address the second objective of this thesis, relationships of interest were investigated within two levels of rural/urban geographic status separately. The relationship between
income inequality and health was significant in rural and urban regions (Table B.3). However, when absolute income measure was included in the linear regression, the independent contribution of the income inequality measure became insignificant in both subgroups (Table B.10).

In testing “absolute income” hypothesis within different rural/urban subgroups, betas for absolute income measure for both rural and urban communities were significant (Table 4.10). Applying the final model in rural and urban subgroups revealed that the betas for absolute income remained significant in rural regions but not in urban regions. The same methodology (section 4.11) for model building applied to establish the best model for rural and urban regions. For urban regions the optimal model was the same but for rural regions variable smoking compared to variable alcohol was better able to explain the variations in the outcome (Figure B.4).

**Rural regions:**

Health Status =77.3+ 0.17 absolute+ (-0.14smoking) + 0.1 education  
Or alternatively, by using standardized coefficients as:

Health Status =0.47 absolute+ (-0.17 smoking) + 0.25 education  

The model explains about 33 % of the variability in the outcome

**Urban regions:**

Health Status =82.4+ 0.08 absolute+ (-0.13 alcohol) + 0.08 education  
Or alternatively, by using standardized coefficients as:

Health Status =0.20 absolute+ (-0.34 alcohol) + 0.23 education  

The model explains about 28 % of the variability in the outcome

Figure B.4. Final Models in Rural and Urban Regions.
B.10 Sensitivity Analysis

All models with the income inequality variable were also tested with all other measures of inequality (section 3.8.8) and no major change in either parameter estimates or R-square observed. We can conclude that the chosen measure was an appropriate measure to capture the rate of income inequality in these data.
B.11 References


Figure B.5: Frequency distribution of measure of population health: Percentage of Canadian Health regions (n=100) that report varying levels of at least good self-rated health, 2004

Figure B.6: Frequency distribution of measure of Absolute Income: Percentage of Canadian Health regions (n=100) that report varying levels of mean absolute personal income, 2004
Figure B.7: Frequency distribution of measure of Income Inequality: Percentage of Canadian Health regions (n=100) that report varying levels of income inequality, 2004

Figure B.8: Frequency distribution of logarithmic transformation measure of Income Inequality: Percentage of varying levels of log of income inequality in Canadian Health regions (n=100), 2004
Figure B.9: Scatter plot of Population Health Status against Income Inequality in Male Regions: X Axis= Log of income inequality, Y Axis= Measure of population health.

Figure B.10: Scatter plot of Population Health Status against Income Inequality in Female Regions: X Axis= Log of income inequality, Y Axis= Measure of population health.
Figure B.11: Scatter plot of Population Health Status against Absolute Income in Male Regions: X Axis= Absolute income level, Y Axis= Measure of population health.

Figure B.12: Scatter plot of Population Health Status against Absolute Income in Female Regions: X Axis= Absolute income level, Y Axis= Measure of population health.
Figure B.13: Scatter plot of Population Health Status against Absolute Income in Senior Regions: X Axis= Absolute income level, Y Axis= Measure of population health.

Figure B.14: Scatter plot of Population Health Status against Absolute Income in Non-Senior Regions: X Axis= Absolute income level, Y Axis= Measure of population health.
<table>
<thead>
<tr>
<th>By Health region</th>
<th>Mean</th>
<th>Median</th>
<th>Min.</th>
<th>Max.</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent reporting good health</td>
<td>87.6</td>
<td>88</td>
<td>80.3</td>
<td>92.5</td>
<td>12.2</td>
</tr>
<tr>
<td>Income Inequality</td>
<td>7.15</td>
<td>4.84</td>
<td>1.47</td>
<td>38.0</td>
<td>36.5</td>
</tr>
<tr>
<td>(Ratio of less than $15,000 to more than $80,000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absolute Income</td>
<td>45.3</td>
<td>46.6</td>
<td>28.3</td>
<td>58.6</td>
<td>30.3</td>
</tr>
<tr>
<td>(Percent with more than $30,000 earners)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent with a regular Family Doctor</td>
<td>86.9</td>
<td>87.6</td>
<td>67.5</td>
<td>97.2</td>
<td>29.7</td>
</tr>
<tr>
<td>Percent of males</td>
<td>49.5</td>
<td>49.5</td>
<td>47.5</td>
<td>52.4</td>
<td>4.90</td>
</tr>
<tr>
<td>Percent older than 65years</td>
<td>15.5</td>
<td>15.6</td>
<td>7.40</td>
<td>23.7</td>
<td>16.3</td>
</tr>
<tr>
<td>Percent with at least secondary education</td>
<td>56.4</td>
<td>56.3</td>
<td>42.4</td>
<td>71.7</td>
<td>29.3</td>
</tr>
<tr>
<td>Percent of daily smokers</td>
<td>18.3</td>
<td>18.6</td>
<td>7.30</td>
<td>27.0</td>
<td>19.7</td>
</tr>
<tr>
<td>Percent of heavy drinkers</td>
<td>21.4</td>
<td>21.6</td>
<td>5.60</td>
<td>43.2</td>
<td>37.6</td>
</tr>
<tr>
<td>Percent of rural population</td>
<td>29.9</td>
<td>29.5</td>
<td>0.00</td>
<td>82.0</td>
<td>82.0</td>
</tr>
<tr>
<td>Percent with strong sense of belonging</td>
<td>65.3</td>
<td>66.6</td>
<td>23.1</td>
<td>95.2</td>
<td>72.1</td>
</tr>
</tbody>
</table>
Table B.2 Pearson Correlation Matrix of Variables

<table>
<thead>
<tr>
<th></th>
<th>Health</th>
<th>Inequality</th>
<th>Absolute</th>
<th>Physician</th>
<th>Sex</th>
<th>Age</th>
<th>Education</th>
<th>Smoking</th>
<th>Alcohol</th>
<th>Rural</th>
<th>Belonging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health</td>
<td>1.00</td>
<td>-0.43§</td>
<td>0.55§</td>
<td>-0.06</td>
<td>0.03</td>
<td>-0.38§</td>
<td>0.50§</td>
<td>-0.39§</td>
<td>-0.32‡</td>
<td>-0.44§</td>
<td>-0.26‡</td>
</tr>
<tr>
<td>Inequality</td>
<td>1.00</td>
<td>-0.77§</td>
<td>0.06</td>
<td>-0.20*</td>
<td>0.32†</td>
<td>-0.37‡</td>
<td>0.17</td>
<td>0.06</td>
<td>0.58§</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>Absolute</td>
<td>1.00</td>
<td>0.01</td>
<td>0.16</td>
<td>-0.50§</td>
<td>-0.25*</td>
<td>-0.20</td>
<td>-0.63§</td>
<td>-0.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physicians</td>
<td>1.00</td>
<td>-0.37‡</td>
<td>0.15</td>
<td>0.04</td>
<td>-0.26‡</td>
<td>-0.02</td>
<td>-0.003</td>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>1.00</td>
<td>-0.38§</td>
<td>-0.34‡</td>
<td>0.30†</td>
<td>0.03</td>
<td>0.18</td>
<td>-0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.00</td>
<td>0.32†</td>
<td>0.04</td>
<td>0.24*</td>
<td>0.37§</td>
<td>0.26‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>1.00</td>
<td>-0.60§</td>
<td>-0.19</td>
<td>-0.65§</td>
<td>0.30†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking</td>
<td>1.00</td>
<td>0.29†</td>
<td>0.38§</td>
<td>-0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol</td>
<td>1.00</td>
<td>0.30†</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p<0.05, †p<0.01, ‡p<0.001, §p<0.0001
Table B.3 Results of linear regression analysis of association between income inequality and health within different levels of covariates (Subgroup Analyses)

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>β</th>
<th>Stand. β</th>
<th>Adj. R²</th>
<th>t*</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income Inequality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>By Absolute Income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rich region</td>
<td>48</td>
<td>-1.29</td>
<td>-0.24</td>
<td>0.04</td>
<td>-1.67</td>
<td>0.1018</td>
</tr>
<tr>
<td>Poor region</td>
<td>52</td>
<td>-1.57</td>
<td>-0.39</td>
<td>0.13</td>
<td>-2.98</td>
<td>0.0044</td>
</tr>
<tr>
<td>By Health Care Utilization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High HCU</td>
<td>48</td>
<td>-2.34</td>
<td>-0.71</td>
<td>0.50</td>
<td>-6.87</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Low HCU</td>
<td>52</td>
<td>-0.16</td>
<td>-0.17</td>
<td>0.01</td>
<td>-1.10</td>
<td>0.2324</td>
</tr>
<tr>
<td>By Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male regions</td>
<td>23</td>
<td>0.14</td>
<td>0.05</td>
<td>-0.05</td>
<td>0.21</td>
<td>0.8381</td>
</tr>
<tr>
<td>Female regions</td>
<td>77</td>
<td>-1.85</td>
<td>-0.52</td>
<td>0.26</td>
<td>-5.26</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>By Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior Regions</td>
<td>39</td>
<td>-1.10</td>
<td>-0.34</td>
<td>0.10</td>
<td>-2.60</td>
<td>0.0530</td>
</tr>
<tr>
<td>Non-Senior regions</td>
<td>61</td>
<td>-1.52</td>
<td>-0.43</td>
<td>0.17</td>
<td>-3.70</td>
<td>0.0005</td>
</tr>
<tr>
<td>By Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High education</td>
<td>46</td>
<td>-0.68</td>
<td>-0.21</td>
<td>0.02</td>
<td>-1.46</td>
<td>0.1515</td>
</tr>
<tr>
<td>Low education</td>
<td>54</td>
<td>-1.10</td>
<td>-0.34</td>
<td>0.10</td>
<td>-2.60</td>
<td>0.0120</td>
</tr>
<tr>
<td>By smoking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High smoking</td>
<td>44</td>
<td>-1.10</td>
<td>-0.34</td>
<td>0.10</td>
<td>-2.35</td>
<td>0.0234</td>
</tr>
<tr>
<td>Low smoking</td>
<td>56</td>
<td>-1.82</td>
<td>-0.53</td>
<td>0.26</td>
<td>-4.55</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>By Alcohol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High alcohol</td>
<td>42</td>
<td>-0.16</td>
<td>-0.37</td>
<td>0.11</td>
<td>-2.49</td>
<td>0.0172</td>
</tr>
<tr>
<td>Low alcohol</td>
<td>58</td>
<td>-1.39</td>
<td>-0.48</td>
<td>0.22</td>
<td>-4.09</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>By Rural/Urban Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>47</td>
<td>-1.14</td>
<td>-0.34</td>
<td>0.09</td>
<td>-2.39</td>
<td>0.0209</td>
</tr>
<tr>
<td>Urban</td>
<td>53</td>
<td>-1.12</td>
<td>-0.30</td>
<td>0.07</td>
<td>-2.19</td>
<td>0.0332</td>
</tr>
<tr>
<td>By Social Capital</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High social capital</td>
<td>46</td>
<td>-1.60</td>
<td>-0.43</td>
<td>0.16</td>
<td>-3.13</td>
<td>0.0031</td>
</tr>
<tr>
<td>Low social capital</td>
<td>54</td>
<td>-1.27</td>
<td>-0.38</td>
<td>0.13</td>
<td>-2.94</td>
<td>0.0049</td>
</tr>
</tbody>
</table>

*  t for income inequality in regression analyses
Table B.4 Pearson coefficient and corresponding p-Values for the correlation between income inequality and health after controlling individually for other covariates

<table>
<thead>
<tr>
<th>Variable Controlled For</th>
<th>Pearson Coefficient</th>
<th>P- Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute</td>
<td>+0.09</td>
<td>0.3816</td>
</tr>
<tr>
<td>Health Care Utilization</td>
<td>-0.43</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Sex</td>
<td>-0.44</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age</td>
<td>-0.33</td>
<td>0.0009</td>
</tr>
<tr>
<td>Education</td>
<td>-0.28</td>
<td>0.0057</td>
</tr>
<tr>
<td>Smoking</td>
<td>-0.40</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Alcohol</td>
<td>-0.41</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Rural</td>
<td>-0.23</td>
<td>0.0241</td>
</tr>
<tr>
<td>Belonging</td>
<td>-0.38</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
Table B.5. Results of linear regression analyses of association between income inequality and health, controlling individually for other covariates

<table>
<thead>
<tr>
<th></th>
<th>$\beta^*$</th>
<th>Stand. $\beta^*$</th>
<th>Adj. $R^2$</th>
<th>$t^*$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inequality</td>
<td>-1.50</td>
<td>-0.43</td>
<td>0.18</td>
<td>-4.76</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Inequality + Absolute</td>
<td>0.50</td>
<td>0.14</td>
<td>0.29</td>
<td>4.11</td>
<td>0.3816</td>
</tr>
<tr>
<td>Inequality + Physicians</td>
<td>-1.50</td>
<td>-0.43</td>
<td>0.17</td>
<td>-4.71</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Inequality + Sex</td>
<td>-1.56</td>
<td>-0.45</td>
<td>0.16</td>
<td>-4.81</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Inequality + Age</td>
<td>-1.15</td>
<td>-0.33</td>
<td>0.22</td>
<td>-3.42</td>
<td>0.0009</td>
</tr>
<tr>
<td>Inequality + Education</td>
<td>-0.91</td>
<td>-0.26</td>
<td>0.30</td>
<td>-2.83</td>
<td>0.0057</td>
</tr>
<tr>
<td>Inequality + Smoking</td>
<td>-1.29</td>
<td>-0.38</td>
<td>0.27</td>
<td>-4.31</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Inequality + Alcohol</td>
<td>-1.36</td>
<td>-0.39</td>
<td>0.24</td>
<td>-4.46</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Inequality + Rural</td>
<td>-0.89</td>
<td>-0.26</td>
<td>0.22</td>
<td>-2.23</td>
<td>0.0241</td>
</tr>
<tr>
<td>Inequality + Social Capital</td>
<td>-1.27</td>
<td>-0.37</td>
<td>0.19</td>
<td>-4.00</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Note. N=100 Canadian Health regions

* Parameter estimates and ts are from the linear regression where Income Inequality was the primary independent variable.
Table B.6 Results of linear regression analysis of association between Absolute income and health within different levels of covariates (Subgroup Analyses)

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>β</th>
<th>Stand. β</th>
<th>Adj. R²</th>
<th>t*</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute Income</td>
<td>100</td>
<td>0.18</td>
<td>0.55</td>
<td>0.30</td>
<td>6.25</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>By Income Inequality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High inequality</td>
<td>48</td>
<td>0.22</td>
<td>0.53</td>
<td>0.27</td>
<td>4.78</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Low inequality</td>
<td>52</td>
<td>0.19</td>
<td>0.44</td>
<td>0.18</td>
<td>3.45</td>
<td>0.0011</td>
</tr>
<tr>
<td>By Health Care Utilization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High HCU</td>
<td>48</td>
<td>0.24</td>
<td>0.76</td>
<td>0.57</td>
<td>7.92</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Low HCU</td>
<td>52</td>
<td>0.12</td>
<td>0.35</td>
<td>0.11</td>
<td>2.64</td>
<td>0.0109</td>
</tr>
<tr>
<td>By Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male regions</td>
<td>23</td>
<td>0.03</td>
<td>0.10</td>
<td>-0.04</td>
<td>0.48</td>
<td>0.6388</td>
</tr>
<tr>
<td>Female regions</td>
<td>77</td>
<td>0.21</td>
<td>0.62</td>
<td>0.38</td>
<td>6.86</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>By Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior Regions</td>
<td>39</td>
<td>0.16</td>
<td>0.45</td>
<td>0.18</td>
<td>3.03</td>
<td>0.0045</td>
</tr>
<tr>
<td>Non-Senior regions</td>
<td>61</td>
<td>0.17</td>
<td>0.52</td>
<td>0.26</td>
<td>4.69</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>By Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High education</td>
<td>46</td>
<td>0.09</td>
<td>0.32</td>
<td>0.08</td>
<td>2.23</td>
<td>0.0306</td>
</tr>
<tr>
<td>Low education</td>
<td>54</td>
<td>0.16</td>
<td>0.50</td>
<td>0.21</td>
<td>3.83</td>
<td>0.0003</td>
</tr>
<tr>
<td>By smoking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High smoking</td>
<td>44</td>
<td>0.14</td>
<td>0.44</td>
<td>0.17</td>
<td>3.16</td>
<td>0.0029</td>
</tr>
<tr>
<td>Low smoking</td>
<td>56</td>
<td>0.20</td>
<td>0.62</td>
<td>0.38</td>
<td>5.88</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>By Alcohol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High alcohol</td>
<td>42</td>
<td>0.22</td>
<td>0.55</td>
<td>0.29</td>
<td>4.22</td>
<td>0.0001</td>
</tr>
<tr>
<td>Low alcohol</td>
<td>58</td>
<td>0.15</td>
<td>0.54</td>
<td>0.28</td>
<td>4.81</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>By Rural/Urban Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>47</td>
<td>0.18</td>
<td>0.52</td>
<td>0.25</td>
<td>4.05</td>
<td>0.0002</td>
</tr>
<tr>
<td>Urban</td>
<td>53</td>
<td>0.14</td>
<td>0.36</td>
<td>0.11</td>
<td>2.76</td>
<td>0.0008</td>
</tr>
<tr>
<td>By Social Capital</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High social capital</td>
<td>46</td>
<td>0.19</td>
<td>0.54</td>
<td>0.28</td>
<td>4.29</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Low social capital</td>
<td>54</td>
<td>0.17</td>
<td>0.51</td>
<td>0.25</td>
<td>4.28</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

* t for absolute income in regression analyses
Table B.7 Pearson coefficient and corresponding p-Values for the correlation between absolute income and health after controlling individually for other covariates

<table>
<thead>
<tr>
<th>Variable Controlled For</th>
<th>Pearson Coefficient</th>
<th>P- Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income inequality</td>
<td>0.39</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Health Care Utilization</td>
<td>0.55</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Sex</td>
<td>0.55</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age</td>
<td>0.45</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Education</td>
<td>0.39</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.51</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Alcohol</td>
<td>0.52</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Rural</td>
<td>0.39</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Belonging</td>
<td>0.51</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
Table B.8. Results of linear regression analyses of association between absolute income and health, controlling individually for other covariates

<table>
<thead>
<tr>
<th></th>
<th>β*</th>
<th>Stand. β*</th>
<th>Adj. R²</th>
<th>t*</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute</td>
<td>0.18</td>
<td>0.55</td>
<td>0.30</td>
<td>6.25</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Absolute + Inequality</td>
<td>0.19</td>
<td>0.56</td>
<td>0.29</td>
<td>4.22</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Absolute + Physicians</td>
<td>0.18</td>
<td>0.55</td>
<td>0.29</td>
<td>6.52</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Absolute + Sex</td>
<td>0.19</td>
<td>0.56</td>
<td>0.29</td>
<td>6.52</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Absolute + Age</td>
<td>0.16</td>
<td>0.48</td>
<td>0.30</td>
<td>4.96</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Absolute + Education</td>
<td>0.13</td>
<td>0.40</td>
<td>0.36</td>
<td>4.21</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Absolute + Smoking</td>
<td>0.16</td>
<td>0.48</td>
<td>0.35</td>
<td>5.78</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Absolute + Alcohol</td>
<td>0.17</td>
<td>0.51</td>
<td>0.34</td>
<td>6.07</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Absolute + Rural</td>
<td>0.15</td>
<td>0.45</td>
<td>0.31</td>
<td>4.16</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Absolute + Social</td>
<td>0.16</td>
<td>0.50</td>
<td>0.30</td>
<td>5.79</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Note. N=100 Canadian Health regions

* Parameter estimates and ts are from the linear regression where Absolute Income was the primary independent variable.
Table B.9. Results of linear regression analyses in different models

<table>
<thead>
<tr>
<th>Models</th>
<th>Independent variable(s)</th>
<th>β</th>
<th>Stand. β</th>
<th>Adj. R²</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 a. Inequality</td>
<td></td>
<td>-1.50*</td>
<td>-0.43</td>
<td>0.18</td>
<td>-4.76</td>
<td>0.0001</td>
</tr>
<tr>
<td>1 b. Absolute Income</td>
<td></td>
<td>0.18†</td>
<td>0.55</td>
<td>0.30</td>
<td>6.25</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>2. All Covariates</td>
<td>(Education)</td>
<td>0.19†</td>
<td>0.54</td>
<td>0.34</td>
<td>6.08</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>(Alcohol)</td>
<td>-0.09†</td>
<td>-0.23</td>
<td></td>
<td>-2.76</td>
<td>0.0100</td>
</tr>
<tr>
<td></td>
<td>(Sex)</td>
<td>0.57†</td>
<td>0.22</td>
<td></td>
<td>2.58</td>
<td>0.0100</td>
</tr>
<tr>
<td>3a. Inequality + Covariates</td>
<td></td>
<td>-0.67†</td>
<td>-0.20</td>
<td>0.34</td>
<td>-1.63</td>
<td>0.1100</td>
</tr>
<tr>
<td>3b. Absolute + Covariates</td>
<td></td>
<td>0.12†</td>
<td>0.37</td>
<td>0.40</td>
<td>3.18</td>
<td>0.0100</td>
</tr>
<tr>
<td>4. Inequality + Absolute + Covariates</td>
<td>(Inequality)</td>
<td>0.47†</td>
<td>0.14</td>
<td>0.39</td>
<td>0.84</td>
<td>0.4100</td>
</tr>
<tr>
<td></td>
<td>(Absolute)</td>
<td>0.15†</td>
<td>0.47</td>
<td></td>
<td>2.81</td>
<td>0.0100</td>
</tr>
<tr>
<td>5a. Inequality + Covariate in model 2</td>
<td></td>
<td>-0.60†</td>
<td>-0.17</td>
<td>0.35</td>
<td>-1.72</td>
<td>0.0900</td>
</tr>
<tr>
<td>5b. Absolute + Covariate in model 2</td>
<td></td>
<td>0.11†</td>
<td>0.32</td>
<td>0.39</td>
<td>3.17</td>
<td>0.0100</td>
</tr>
<tr>
<td>6a. Absolute + Alcohol+ Education</td>
<td></td>
<td>0.12†</td>
<td>0.37</td>
<td>0.39</td>
<td>3.99</td>
<td>0.0001</td>
</tr>
<tr>
<td>6b. Absolute + Alcohol+ Smoking +Belonging</td>
<td></td>
<td>0.13†</td>
<td>0.40</td>
<td>0.39</td>
<td>4.73</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>6c. Absolute + Alcohol+ Smoking +Education</td>
<td></td>
<td>0.13†</td>
<td>0.38</td>
<td>0.39</td>
<td>4.10</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Models</td>
<td>β</td>
<td>Stand. β</td>
<td>Adj. R²</td>
<td>t</td>
<td>p</td>
<td></td>
</tr>
<tr>
<td>------------------------------</td>
<td>-------</td>
<td>----------</td>
<td>---------</td>
<td>-------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>6d.Absolute+Alcohol+Smoking +Physicians+belonging</td>
<td>0.13†</td>
<td>0.40</td>
<td>0.40</td>
<td>4.73</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>6e.Absolute+Alcohol+Smoking +Education+Belonging</td>
<td>0.12†</td>
<td>0.35</td>
<td>0.40</td>
<td>3.84</td>
<td>0.0002</td>
<td></td>
</tr>
</tbody>
</table>

* β, t, and p for Income Inequality variable.
† β, t, and p for Absolute Income variable.
§ β, t, and p for individual covariates.
Note. MSE for all models in 6 was between 3.19 and 3.3.
Table B.10 Results of linear regression analysis of association between income inequality, absolute income and health within different levels of rural/urban status (Cut off=30)

<table>
<thead>
<tr>
<th>A) Independent variable= Income Inequality</th>
<th>n</th>
<th>β</th>
<th>Standardized β</th>
<th>Adjusted R²</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural regions</td>
<td>47</td>
<td>-1.14</td>
<td>-0.34</td>
<td>0.09</td>
<td>-2.39</td>
<td>0.0209</td>
</tr>
<tr>
<td>Urban regions</td>
<td>53</td>
<td>-1.12</td>
<td>-0.29</td>
<td>0.07</td>
<td>-2.19</td>
<td>0.0332</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B) Independent variable= Absolute Income</th>
<th>n</th>
<th>β</th>
<th>Standardized β</th>
<th>Adjusted R²</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural regions</td>
<td>47</td>
<td>0.18</td>
<td>0.52</td>
<td>0.25</td>
<td>4.05</td>
<td>0.0002</td>
</tr>
<tr>
<td>Urban regions</td>
<td>53</td>
<td>0.14</td>
<td>0.36</td>
<td>0.11</td>
<td>2.76</td>
<td>0.0080</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C) Independent variables= Income Inequality+ Absolute Income</th>
<th>n</th>
<th>β</th>
<th>Standardized β</th>
<th>Adjusted R²</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural regions</td>
<td>47</td>
<td>0.57</td>
<td>0.17</td>
<td>0.25</td>
<td>0.83</td>
<td>0.4119</td>
</tr>
<tr>
<td>Urban regions</td>
<td>53</td>
<td>0.22</td>
<td>0.06</td>
<td>0.10</td>
<td>0.23</td>
<td>0.8213</td>
</tr>
</tbody>
</table>

* For income inequality variable
Appendix C: Ethics Approval

QUEEN'S UNIVERSITY HEALTH SCIENCES & AFFILIATED TEACHING
HOSPITALS RESEARCH ETHICS BOARD

November 2, 2007

This Ethics Application was subject to:

☑ Full Board Review
Meeting Date:
☑ Expedited Review

Mr. Afshin Vafaei
Department of Community Health and Epidemiology
Queen's University

Dear Mr. Vafaei,

Study Title: Relationships between income inequality and health: a study at national and provincial levels
Co-Investigators: Dr. W. Pickett and Dr. M. Rosenberg

I am writing to acknowledge receipt of your recent ethics submission. We have examined the protocol for your project (as stated above) and consider it to be ethically acceptable. This approval is valid for one year from the date of the Chair's signature below. This approval will be reported to the Research Ethics Board. Please attend carefully to the following list of ethics requirements you must fulfill over the course of your study:

➤ Reporting of Amendments: If there are any changes to your study (e.g. consent, protocol, study procedures, etc.), you must submit an amendment to the Research Ethics Board for approval. (see http://www.queensu.ca/ports/eth.htm).

➤ Reporting of Serious Adverse Events: Any unexpected serious adverse event occurring locally must be reported within 2 working days or earlier if required by the study sponsor. All other serious adverse events must be reported within 15 days after becoming aware of the information.

➤ Reporting of Complaints: Any complaints made by participants or persons acting on behalf of participants must be reported to the Research Ethics Board within 7 days of becoming aware of the complaint. Note: All documents supplied to participants must have the contact information for the Research Ethics Board.

➤ Annual Renewal: Prior to the expiration of your approval (which is one year from the date of the Chair's signature below), you will be reminded to submit your renewal form along with any new changes or amendments you wish to make to your study. If there have been no major changes to your protocol, your approval may be renewed for another year.

Yours sincerely,

[Signature]
Chair, Research Ethics Board

[Signature]
Date

Study Code: EPID-252-07

➤ Investigators please note that if your trial is registered by the sponsor, you must take responsibility to ensure that the registration information is accurate and complete.
QUEEN'S UNIVERSITY HEALTH SCIENCES & AFFILIATED TEACHING
HOSPITALS RESEARCH ETHICS BOARD

The membership of this Research Ethics Board complies with the membership requirements for Research Ethics Boards as defined by the Tri-Council Policy Statement; Part C Division 5 of the Food and Drug Regulations, OHRP, and U.S DHHS Code of Federal Regulations Title 45, Part 46 and carries out its functions in a manner consistent with Good Clinical Practices.

Federalwide Assurance Number : #FWA00004184
#IRB00001173

Current membership of the Queen's University Health Sciences & Affiliated Teaching Hospitals Research Ethics Board

Dr. A.F. Clark  Emeritus Professor, Department of Biochemistry, Faculty of Health Sciences, Queen's University (Chair)

Dr. H. Abdollah  Professor, Department of Medicine, Queen's University

Rev. T. Deline  Community Member

Dr. M. Evans  Community Member

Prof. L. Keeping-Burke  Assistant Professor, School of Nursing, Queen's University

Dr. J. Low  Emeritus Professor, Department of Obstetrics and Gynaecology, Queen's University and Kingston General Hospital

Dr. W. Racz  Emeritus Professor, Department of Pharmacology & Toxicology, Queen’s University

Dr. H. Richardson  Assistant Professor, Department of Community Health & Epidemiology Project Coordinator, NCIC CTG, Queen's University

Dr. B. Simchison  Assistant Professor, Department of Anesthesiology, Queen's University

Dr. A.N. Singh  WHO Professor in Psychosomatic Medicine and Psychopharmacology Professor of Psychiatry and Pharmacology Chair and Head, Division of Psychopharmacology, Queen’s University Director & Chief of Psychiatry, Academic Unit, Quinte Health Care, Belleville General Hospital

Dr. E. Tsai  Assistant Professor, Department of Paediatrics and Office of Bioethics, Queen's University

Rev. J. Warren  Community Member

Ms. K. Weisbaum  LL.B. and Adjunct Instructor, Department of Family Medicine (Bioethics)

Ms. S. Wood  Associate Director, Office of Research Services (Ex-Officio)