THE RELATIONSHIP BETWEEN INCOME INEQUALITY, WELFARE REGIMES AND AGGREGATE HEALTH

By

KI-TAE KIM

A Thesis submitted to The University of Birmingham

For the degree of

DOCTOR OF PHILOSOPHY

Health Service Management Centre

College of Social Sciences

University of Birmingham

August 2016

ABSTRACT

The Scandinavian welfare regime is expected to have better aggregate health than other welfare regimes due mainly to its narrow income inequality. This theoretical expectation is in part related to the Wilkinson Hypothesis that, in industrialised nations, a society's narrow income inequality enhances its aggregate health. This thesis tests both of the above propositions. This is achieved by means of four methods not previously applied to this field, namely a 'review of reviews', a decomposition systematic review, a new case selection method, and a use of the OECD regional dataset for the cross-national comparative health study.

These new methodological approaches lead to four main findings. First, the Scandinavian welfare regime shows worse-than-expected aggregate health outcomes. This thesis terms this counterintuitive finding as 'the second Scandinavian puzzle'. Second, the East Asian welfare regime shows unexpectedly good aggregate health, which is proposed as 'the East Asian puzzle'. Third, regarding the Wilkinson Hypothesis, it is income, rather than income inequality, which is a statistically significant determinant of aggregate health. Fourth, the effects on health of income inequality or welfare regimes reverse over a certain threshold of age, which is termed here 'the age threshold effect'.

ACKNOWLEDGEMENTS

"Can you write something on your study and submit it every week?"

Professor Martin Powell, my supervisor, asked in a meeting in September, 2012 when I first came to Birmingham. That was how I began to sit up all night every Friday morning since then. At first, I submitted a 305-word note. I thought that he must be very busy and I had to minimise the length for him to read it all. He said, however, he wanted a longer writing. Next Friday morning, I sent him a longer 2300-word research note. A few hours later, he sent me back the document with nine comments added and with the following sentence at the end.

"This is definitely the style/ length that you need to adopt … However, we now need to start narrowing down your area of interest, and start thinking about focus of work and RQ (research questions). For what it is worth, my idea (but it is YOUR thesis) is that there is a gap involving EA regimes, and the least studies features (PP mix; publicness; stratification; include health inequalities)"

Since then, I saw every rising sun on Friday with a bit pinkish eyes until the end of the year to meet his high standard. If this thesis makes a contribution, I owe it to my supervisor. I can never forget joyful and meaningful conversations with him at pubs and beer festivals as well. I am not sure if I can ever become such a respected supervisor to any PhD student as he is to me.

Dr. Iestyn Williams, my vice supervisor, has supported me throughout with his supervision, humour and intelligence. He has special talents of understanding my sometimes broken, verbal English. Without his insightful comments, the quality of my thesis would be much more downgraded. I don't know how to express my gratitude to my amazing supervisors.

I also thank two 'unofficial' supervisors. Dr. Ross Millar, my roommate at Office 49 at Park House for one and half years, gave me precious advice not only on my thesis but also about life in the UK. He, always a big smile on his face, has been good company with his wisdom and cheerfulness. Thanks to Professor Baek Seung-ho at Catholic University in Korea, I was able to learn statistical methods for this thesis. My thesis is hugely indebted to him. He also kindly draws Figures 6-3~7 in my thesis.

I also thank many friends at Birmingham University such as the wonderful librarians at HSMC (Wendy Spurr, Lucy Drake and Rachel Posner) and my fellow PhD students and researchers (Dr. Lee Euncheong, Dr. Lim Cheol-eon, Sara-Jane Fenton, Dr. Alexis Paton, Dr. Robin Miller). I also appreciate Mr. Ju Sang Sun and his wonderful family for their help while my family first settled in Birmingham. Rev. Samson Park generously becomes my tennis partner every Thursday morning.

My family have given me huge emotional support from start to finish. Love to my parents and parents-in-law, three sisters-in-law, two brothers, two brothers-in-law, and their adorable boys and girls including Chaewon, Seohan, Jisun, Buhm Joon, Soohan, and Soohyun. My two sons, Eehan and Eejoon, born in 2012 and 2015, are simply sunshine to me throughout the tough days. Jungsun Min, my wife, has struggled to raise the two boys and supports her often grumpy husband for the last four years. Without her wisdom and brightness, I would not be able to write this thesis. I dedicate my thesis to my wife.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1 Introduction

Gøsta Esping-Andersen's (1990) 'The Three Worlds of Welfare Capitalism' has been acclaimed as "one of the most cited works in social policy" (Powell and Barrientos, 2015, p. 241), "a classic" (Emmenegger, Kvist, Marx, and Petersen, 2015, p. 3) or "paradigmatic" (Kersbergen and Vis, 2015, p. 112). In the typology of 18 traditional welfare states, Esping-Andersen (1990; 1999a) suggests three welfare regimes of Liberal, Conservative and Social Democratic 'worlds'. Among the three types, his preference for the Social Democratic model seems evident from the beginning. For example, he notes that "the beauty of the social democratic strategy was that social policy would also result in power mobilization… parliamentary class mobilization is a means for the realization of the socialist ideals of equality, justice, freedom, and solidarity" (Esping-Andersen, 1990, p. 12).

Similarly, other researchers acclaim the Social Democratic model with its typically universal welfare states exerting a positive influence on the quality of the people's lives. In comparison with the other welfare regimes, for example, the egalitarian welfare regime turns out to have the positive empirical outcomes in terms of income inequalities (e.g. Coburn, 2004), social capital (e.g. Kääriäinen and Lehtonen, 2006), intergenerational mobility (e.g. Esping-Andersen, 2014), infant mortality (e.g. Chung and Muntaner, 2007; Raphael, 2013), poverty reduction (e.g. Fouarge and Layte, 2005; Whelan and Maitre, 2010) and life satisfaction (e.g. Niedzwiedz, Katikireddi, Pell, and Mitchell, 2014). Goodin, Headey, Muffels, and Dirven (1999) provide a comprehensive approach by comparing the panel data of three key welfare states of US, Germany and the Netherlands, respectively representing the Liberal, Conservative and Social Democratic models. They conclude that the Social Democratic welfare regime shows the best results in most of their six indicators: promoting efficiency, reducing poverty, promoting equality, promoting integration, promoting stability and promoting autonomy.

Within the Social Democratic welfare regime, the Scandinavian welfare states are generally regarded as its core or prototype for its characteristic egalitarian welfare states (Kvist, Fritzell, Hvinden, and Kangas, 2012). In Esping-Andersen (1990), the Social Democratic welfare regime consists of not only three Scandinavian states of Sweden, Norway, Denmark but also three other continental European states: Belgium, Austria and the Netherlands (p. 52).

However, he often uses the three Scandinavian states as the representatives of the Social Democratic welfare regime using the expressions like “the Scandinavian social democratic welfare states” (p. 87) or “Scandinavian model” (p. 156). Consequently, Esping-Andersen (1999a) categorises Denmark, Sweden, Finland and Norway as the four nations forming the Social Democratic welfare regime (p. 77).

In many subsequent comparative studies, the three to five Scandinavian welfare states, with or without Finland and Iceland, are considered as equivalent to the Social Democratic welfare regime (e.g. Siaroff, 1994; Bonoli, 1997; Obinger and Wagschal, 2001; Shalev, 2007; Vrooman, 2012). In the recent article 'welfare regime and social stratification' marking the 25th anniversary of his book's publication, Esping-Andersen (2015) uses the Scandinavian model as synonymous with the Social Democratic welfare regime and is still in praise of the type: "the Scandinavian welfare regime has been, comparatively speaking, substantially more effective in equalizing the opportunity structure" (p.132). For the five North European states, various terms have been used such as Nordic, Scandinavian and Social Democratic. To avoid confusion, this thesis consistently uses the term 'the Scandinavian welfare regime' for these five states unless a specific term is directly quoted from other authors.

Along with social policy writers, comparative health researchers also expect the Social Democratic welfare regime, or particularly the Scandinavian states to have the best aggregate health (e.g. Mackenbach, 2012; Huijts and Eikemo, 2009; Richter et al., 2012). One of the main reasons behind this expectation is that the Scandinavian welfare states have the narrowest disposable income inequalities due to their generous welfare states. For example, it is stated that "population health is enhanced by the relatively generous and universal welfare provision of the Scandinavian countries" (Bambra, 2011, p. 2), "The health effect of welfare provision is more or less accepted" (Mackenbach, 2011, p. 1) or "the universalistic and redistributive approach taken in the Nordic countries makes positive overall health outcomes" (Huijts and Eikemo, 2009, p. 2).

This expectation is closely related to 'the Wilkinson Hypothesis' that in a society over a certain threshold of GDP per capita, narrow income distribution causes a positive effect on aggregate health (Wilkinson, 1992, 1996; Wilkinson and Pickett, 2009). Although not adopting the welfare regime typology, Wilkinson and colleagues also present some

Scandinavian nations as the healthiest when compared with populations in different types of welfare states. Some contend that the relation between narrow income inequality and better aggregate health is a logical consequence because of the theoretical curvilinear relation between income inequality and individual health (Gravelle, 1998; Jen, Jones, and Johnston, 2009a).

To simplify their argument, a society can be assumed to consist of only two persons: one poor individual and the other rich. The same 100 pounds in additional income would be significantly beneficial for a poor person's health, but the benefit would be almost negligible for the wealthy person. If the rich individual gives 100 pounds to the poor person (i.e. income inequality is narrowed between the two), this logically leads to better aggregate health because the benefit of the poor person's health would outstrip the relatively minimal damage to the rich person. This simple logic can be applied to a society of millions of people. It follows that in the Scandinavian states where the ratio of the poor population is the lowest (in other words, where the rich share more wealth with the poor than any other societies) the detrimental effects of poverty would be lowest, which will enhance the average level of health in the Scandinavian welfare regime. The logic behind the theoretical expectation by many researchers is similar to a syllogism as follows. The Wilkinson Hypothesis is placed in the middle as the proposition II.

I) Scandinavian welfare states (A) have relatively low income inequalities (B): A => B

II) Narrow income inequality (B) would lead to better aggregate health (C): B => C

III) Scandinavian welfare states (A) have better aggregate health (C): A => C.

Some cross-national health researchers reach conclusion III based on this plausible reasoning. The Scandinavian universal welfare states have the lowest disposable or after-tax income inequalities, which leads to better aggregate health outcomes.

However, given the logic of the syllogism, if proposition I or II turns out to be erroneous, the conclusion has to be called into question. First of all, in the case of the Scandinavian income inequality levels (I), the majority of cross-national studies affirm that Scandinavian nations have long succeeded in maintaining a fair disposable income distribution (e.g. Gottschalk and Smeeding, 2000; OECD, 2011). However, the proposition II or the Wilkinson Hypothesis (Wilkinson, 1992, 1996; Avendano and Hessel, 2015) is debatable. Scores of studies have been published for decades, carrying evidence to support the hypothesis (e.g. Rodgers, 1979; Beckfield, 2004; Pickett and Wilkinson, 2015), but a similar number of studies have also been published to refute the hypothesis (e.g. Judge, 1995; Pop, Ingen, and Oorschot, 2012).

Skeptical commentators on the hypothesis suggest that the relationship between income inequality and aggregate health vanishes after other health determinants are controlled for, such as GDP per capita (Mello and Milyo, 2001) or educational attainment (Muller, 2002). Observing the growing evidence in this critical perspective, a British Medical Journal (BMJ) editorial declares that evidence in support of the hypothesis "has disappeared" (Mackenbach, 2002, p. 1). Avendano and Hessel (2015) also state that "there is no strong evidence" (p. 597) for the hypothesis. Two review articles reach contrasting conclusions on the hypothesis with Lynch et al. (2004) reaching a negative conclusion and Wilkinson and Pickett (2006) giving a positive verdict on the hypothesis.

Overall, findings in either empirical or review studies on the hypothesis can be seen as quite inconsistent or even contradictory.

With the lack of a consensus on the Wilkinson Hypothesis, the expectation for Scandinavia's good aggregate health, based in part on the hypothesis, requires testing. In fact, among some empirical studies examining the relationship between welfare regimes and aggregate health, the conclusions are also largely inconsistent. Some argue that the Scandinavian or Social Democratic welfare regime outperform the other welfare regimes in terms of aggregate health indicators such as infant mortality rate (e.g. Bambra, 2006a; Raphel, 2013), while other studies refute the relationship (e.g. Kangas, 2010; Karim et al., 2010; Regidor, 2011). In the latest systematic review on the relation between welfare regimes and aggregate health, Bergqvist et al. (2013) conclude that "results are diverse and contradictory" (p. 1234).

Given the inconsistent empirical findings, the expectation of Scandinavia's good aggregate health in comparative perspectives requires further analysis in order to reach a consensus. However, many comparative health researchers seem to rather uncritically accept the hypothesis. They state that Scandinavian nations have "enviable health profiles" (Raphael, 2014, p. 10), "the highest level of population health" (Richter et al., 2012, p. 860), are "doing well in overall health outcomes" (Bambra, 2013, p. 713), "rank higher on various population health indicators than the other regimes" (Hurrelmann, Rathmann and Richter, 2010, p. 6) and there are "better health outcomes … for social democratic welfare states" (Chung and Muntaner, 2008, p. 282).

Wilkinson and Pickett (2009) also state: “Internationally, at the healthy end of the distribution, we always seem to find the Scandinavian countries…” (p. 172). This is the main focus of this thesis: the mismatch between the prevalent expectation for, or belief in, Scandinavia’s good aggregate health and the empirically inconsistent findings on the Scandinavian health. In other words, researchers may have been reluctant to acknowledge some counterintuitive findings regarding Scandinavian health. This reluctance is surprising given the significant attention paid to another counterintuitive finding on a different aspect of Scandinavian health: its unexpected underperformance in narrowing health inequalities. The term ‘underperformance’ is used because the Scandinavian states with the lowest income inequalities are theoretically expected to show the narrowest health inequalities within their population. However, empirical findings suggest that the Scandinavian welfare regime does not show the narrowest health inequalities in comparison with other welfare regimes (e.g. Mackenbach, Kunst, Cavelaars, Groenhof, and Geurts, 1997).

Researchers actively recognise this counterintuitive evidence and name it as a “paradoxical finding” (Huijts and Eikemo, 2009, p. 452), a “puzzle” (Bambra, 2011, p. 740), a “puzzling finding” (Lahelma and Lundberg, 2009), a “Scandinavian paradox” (Richter et al., 2012a, p. 860; Hurrelmann, Rathmann and Richter, 2010, p. 6), which is “difficult to digest” (Mackenbach, 2011, p. 1). Several journal articles explore the cause of the challenging findings (e.g. Huijts and Eikemo, 2009; Bambra, 2011). For them, Scandinavian universal welfare states and their relatively fair labour markets should have led to narrow health inequalities in comparison with other welfare states.

Given the intense and widespread interest in Scandinavia’s unexpected underperformance in health inequalities, it is noteworthy that they presuppose

Scandinavia’s good aggregate health and overlook the empirical findings revealing the Scandinavia’s worse-than-expected aggregate health. It is in general the same group of researchers who acknowledge Scandinavia’s unexpectedly wide health inequalities and at the same time uncritically assume Scandinavia’s good aggregate health. For example, they state that while the Scandinavian welfare state “makes positive overall health outcomes, it does not sufficiently reduce the relative inequalities in health” (Huijts and Eikemo, 2009, p. 2) or “in contrast to their comparatively strong performance in terms of overall health, … the Scandinavian welfare states do not have the smallest health inequalities” (Bambra, 2011, p. 740).

This is the gap that this thesis focuses on: the somewhat overlooked aspect of Scandinavia’s worse-than-expected aggregate health. It proposes this counterintuitive finding as ‘the second Scandinavian puzzle’. The term ‘second’ is added because it needs to be distinguished from the first puzzle of Scandinavia’s underperformance in narrowing health inequalities. Table 1-1 presents the differences between the first and second Scandinavian puzzles and demonstrates the gap unexplored as seen on the bottom right cell.

A secondary focus of this thesis is on aggregate health in the largely neglected East Asian welfare regime (but see Chuang, Chuang, Chen, Shi, and Yang, 2012; Popham, Dibben, and Bambra, 2013). The East Asian welfare regime needs to be examined as a distinctive unit, separate from the Western welfare regimes for the two reasons.

Table 1-1. The First and Second Scandinavian Puzzle?

	The First Scandinavian Puzzle	The Second Scandinavian Puzzle
What is it about?	Scandinavia's unexpectedly wide health inequalities	Scandinavia's unexpectedly low aggregate health
Evidence supporting the argument	Mackenbach et al. (2008), Espelt et al. (2008), etc.	Kangas (2010), Karim et al.(2010), Regidor (2011), etc.
Evidence denying the argument	Popham et al. (2013), etc.	Bambra (2006a), Raphel (2013), etc.
The nicknames of the puzzle? (Who acknowledge the puzzle?)	"puzzling finding" (Lahelma & Lundberg, 2009, p. 452), "Scandinavian paradox" (Richter et al., 2012a, p. 860; Hurrelmann, Rathmann & Richter, 2010, p. 6), "puzzle" (Bambra, 2011, p. 740).	?

First, Japan is the single non-Western nation examined by Esping-Andersen (1990) and could not form any distinctive welfare regime on its own. However, South Korea has emerged as one of the new welfare states (e.g. Powell and Kim, 2014) and the East Asian welfare model has begun to draw attention from welfare state researchers for its distinctiveness (Kwon, 1997; Wilding, 2008; Hudson and Hwang, 2013). For example, "this coincidence of welfare expansion with economic liberalzation, a shift towards a more post-industrial economy and more pluralist polity, is highly significant and cannot easily fit within the classic 'Western' theories of welfare development" (Hudson and

Hwang, 2013, p. 36). Consequently, these two East Asian welfare states, sometimes with other neighbouring states such as Taiwan or Singapore, are regarded as forming the distinct East Asian welfare regime (e.g. Lee and Ku, 2007; Walker and Wong, 2005).

Second, the East Asian welfare regime shows remarkable health outcomes. Japan has kept the reputation for its best longevity records for the last three decades (OECD, 2013a). South Korea, the only East Asian OECD member state except for Japan, has a life expectancy of 80.0 years in 2009, leveling with that of Finland (OECD, 2013a). South Korea increased its life expectancy by 27.9 years for the last 50 years. This compares well to the OECD average increase in life expectancy for the period, 11.2 years. Given these impressive health outcomes, the East Asian regime warrants further examination as a separate entity in this between-regime comparative health research.

1.2. Research Questions

This thesis aims to introduce and investigate 'the second Scandinavian puzzle' and to consider its implications for theory, research and policy. To this end, this thesis sets the two research questions:

1) What is the relationship between income inequality and aggregate health?
(i.e. the Wilkinson Hypothesis)
2) What is the relationship between welfare regimes and aggregate health?
(related to the second Scandinavian puzzle and the East Asian puzzle)

There are three reasons to seek answers to the first research question. First, as discussed, this question is closely associated with the second research question. In case the relationship between income inequality and aggregate health proves to be insignificant, the logical consequence is that Scandinavian states have fewer reasons to have better aggregate health than other welfare sates. Before the second Scandinavian puzzle is examined, it is a prerequisite to seek answers to the first research question.

Second, empirical findings on this hypothesis are inconsistent over the past decades and what researchers have found so far have been inconclusive (e.g. Wilkinson, 1992; Judge, 1995; Pop, Ingen, and Oorschot, 2012). This thesis uses refined statistical techniques and newly introduced datasets to enable us to investigate the controversial relationship from clearer perspectives.

Third, a focus on income inequalities is important, since they have continued to widen in industrialised nations for the last decades, and are claimed to be at their worst since the Great Depression (Reich, 2011; OECD, 2011). The time-series data over the last decades may show if and how the deepening income distribution worsens aggregate health in the wealthy societies.

The second research question also warrants examination for the following three reasons. First, it is only relatively recently that researchers have begun to examine this relationship between welfare characteristics and aggregate health at national or welfare regime levels (e.g. Chung and Muntaner, 2007; Karim et al., 2010; Tapia Granados, 2010). Their general expectation is that the Scandinavian or Social Democratic welfare regime would record the best aggregate health, but their conclusions fail to reach a consensus. This thesis proposes the “diverse and contradictory” (Bergqvist et al., 2013,

p. 1234) findings as ‘the second Scandinavian puzzle’ and attempts to seek the answers.

Second, as discussed, many comparative health researchers appear to disregard counterintuitive findings on Scandinavia’s relatively low level of aggregate health and arguably accept too readily Scandinavian excellence in aggregate health (e.g. Richter et al., 2012; Bambra, 2013). This stance might be claimed to be based on *a priori* reasoning or ‘Swedocentrism’ (Abrahamson and Wehner, 2006, p. 3) rather than on accumulated knowledge. This needs to be tested and examined.

Third, the debated relationship also raises a question regarding the role of the welfare state. As Bambra (2006a, p. 53) notes, the research on the relationship is to “examine what welfare states actually do rather than how much they are afforded or which services they provide”. If the most egalitarian and universal welfare model nested in the Scandinavian regions turns out to underperform in enhancing aggregate health - one of the most important indicators to gauge people’s quality of lives - people need to ask again what benefits the acclaimed welfare model has been able or unable to carry. The answers to these daunting questions may have critical implications for the studies on and policies of welfare states.

1.3. Definition of Key Concepts

In addition to clarifying the research questions, clear definitions of the three key words within the research questions are needed: aggregate health, income inequality and welfare regime.

The term ‘aggregate health’ is to indicate an average level of health in region-, nation- or welfare regime-level population. This researcher distinguishes this from other similar terms such as ‘population health’ or ‘public health’ and avoid the use of the latter concepts. In an article devoted to population health terminology, Kindig and Stoddart (2003) define ‘population health’ as “the health outcomes of a group of individual, including the distribution of such outcomes within the group” (p. 380), admitting that the definition also encompasses “health inequality and inequity” (p. 381). This thesis avoids this term to clarify that the focus here is solely on aggregate health rather than health inequalities. In addition, the term ‘public health’ is avoided in this thesis because it is defined as “What we do as a society collectively to assure conditions in which people can be healthy” (IOM, 1998; cited by Kindig, 2007, p. 146). However, when this thesis directly quotes statements from other literature, the term ‘population health’ may mean ‘aggregate health’ because the two words are often used interchangeably.

Second, income inequality, expressed in “the most commonly used measure” (Kondo et al., 2009, p. 2) of Gini coefficient in this thesis, is rather technically defined as “half of the arithmetic average of the absolute differences between all pairs of incomes in a population, the total then being normalised on mean income”(Kawachi, Subramanian, and Almeida-Filho, 2002, p. 649). According to the calculation, if every individual or household earns completely equal income, the Gini coefficient will be zero. If a single individual or household takes all income, it will be 1.0. The more equal a society is, the closer to zero the coefficient is. Among various income indicators, most of comparative health researchers use disposable (net or post-tax) income indicators because it captures the actual amount money people earn in the end (e.g. Wilkinson, 1992; Lynch and Kaplan, 1997; Mackenbach et al., 2008; Wilkinson and Pickett, 2009). However, this thesis uses market (gross or taxable) income in addition to the disposable income to analyse effects of 'tax and transfer' that is added to, or deducted from, gross income to form disposable income. The effect of market income inequality is also analysed.

Third, the term 'welfare regime' means "the combined, interdependent way in which welfare is produced and allocated between state, market, and family" (Esping-Andersen, 1999a, p. 35). This thesis relies on this classic definition but does not adopt Esping-Andersen's (1990, 1999a) tripartite categorisation of welfare regimes: Liberal, Conservative and Social Democratic. The reason for this is that that the typology, based on data in 1980s, has limitations in covering the emerging welfare states (see Kim, 2015) in Southern Europe (e.g. Powell and Barrientos, 2004), East European (e.g. Fenger, 2007) and East Asia (e.g. Powell and Kim, 2014).

In addition, it should be noted that this thesis uses the term 'Scandinavian welfare regime' rather than the Social Democratic or Nordic welfare regime unless the latter terms appear in direct quotations. There are three reasons for this. First, the member states of the Social Democratic welfare regime differ in almost every empirical study (e.g. Esping-Andersen, 1990; Bonoli, 1997; Ferragina, Seeleib-Kaiser, and Tomlinson, 2013) and the term itself may cause confusion rather than clarify the boundary of the concept. Second, the traditionally and culturally homogenous Scandinavian nations, at least in relative terms, are often regarded as forming a distinctive welfare regime and, sometimes, identical to the Social Democratic welfare regime itself (e.g. Shalev, 1996; Bambra, 2004; Vrooman, 2012). Third, the Scandinavian welfare regime has already been adopted as a more specific term in contrast to Liberal or Conservative welfare regime (e.g. Esping-Andersen, 2015).

Consequently, this thesis uses the term of the Scandinavian welfare regime to clearly specify its focus as well as avoid any conceptual confusion. The Scandinavian welfare regime denotes the five regional states: Sweden, Finland, Norway, Denmark and Iceland. On the case of Iceland, its categorization can be debatable, but this thesis classifies it into the Scandinavian welfare regime (see Chapter 4 for more details).

Another issue regarding the concept of the Scandinavian welfare regime is the reliability of the concept over the past decades. Analysing the case of Denmark, Kvist and Greve (2011) observe that the Scandinavian state has increasingly resorted to occupational and fiscal welfare measures, which transformed the welfare model into a multi-tiered welfare state. Danforth (2014) claim that the distinct Social Democratic welfare regime, once salient between 1970s and 1990s, no longer exists, after examining the chronological changes of 18 traditional welfare states. He argues that the Social Democratic model, merged with the Conservative model, has formed a 'European' model since 2000. Moreover, the Scandinavian welfare states overall have seen their income inequalities widen since the 1990s (Fritzell, Bäckman, and Ritakallio, 2012).

However, other empirical findings support the persistence of the Scandinavian model. Based on a cluster analysis for the pooled data over the relatively recent period of 2005-2012 with its focus solely on 14 EU member states and welfare state income indicators encompassing both old social risks and new social risks, Ferragina, Seeleib-Kaiser, & Spreckelsen (2015) note that the four worlds of welfare capitalism including Conservative, Liberal, Mediterranean and Social democratic, remain salient. In terms of old social risks indicators such as unemployment rates or replacement rates for pensioners, five nations of Belgium, Denmark, Sweden, Austria and Finland are grouped together. In terms of other criteria of new social risks indicators such as child and youth poverty rates, percentage in youth in education and female labour force participation, the five states of Denmark, Sweden, Germany, Netherlands and Finland are grouped together. In any perspectives, the three Scandinavian nations (Sweden, Denmark and Finland) are clustered in the same group while Norway and Iceland were not included in the analysis. The two rounds of cluster analysis "replicate very closely the theoretical four-cluster typology" (Ferragina, Seeleib-Kaiser, & Spreckelsen, 2015, p. 287) affirming the distinctive status of the Scandinavian welfare regime.

This finding is also consistent with the observations such as "These profound changes have taken place in such a way that although core characteristics are still in place, new structures and understandings of the welfare state are also developing" (Kvist and Greve, 2010, p. 146). In addition, the majority of empirical studies assert the distinctiveness of the Scandinavian model and place the Scandinavian welfare states into a distinctive category (e.g. Ferrera, 1996; Schröder, 2009;; Vrooman, 2012; Ferragina, Seeleib-Kaiser & Tomlinson, 2013). (See more details on the categorisation of the five Scandinavian states in Chapter 5 and more details on debates on the changes of Scandinavian welfare state model over decades in Chapter 8)

Regarding the distinct East Asian welfare regime consisting of Japan and South Korea, especially in comparing its combined health outcomes with those of other welfare regimes, the following four questions can be raised. First, some have called into question East Asian distinctiveness (e.g. Esping-Andersen, 1997; Kim, 2011). Second, it is also debatable whether the welfare states in South Korea and Japan share enough in common to be categorised as a welfare regime (e.g. Wilding, 2008). In other words, the first question relates to distinctiveness *across* regimes (i.e. comparing the proposed new East Asian regime with established Western versions). The second questions relates to similarity of the two countries *within* the proposed East Asian regime. Third, a distinctive welfare regime with only two national cases may be problematic, when other welfare regimes consist of up to seven welfare states. Fourth, including another candidate nation such as Taiwan or Singapore in the East Asian welfare regime raises the issue of case selection (Kim, 2015) and could change the findings of this thesis.

With regard to the first question, Esping-Andersen (1997) states that Japan is either a hybrid between conservative and liberal model or its welfare state has not yet sunk its roots. For him, "the kind of welfare-uniqueness that authors such as Ezra Vogel (1973) stress is hardly at all unique to Japan" (p. 187). Kim (2011) also concludes that East Asian 'exceptionality' in relation to the Western welfare model is gradually fading after he reviews changes in social policy in four East Asian nations of South Korea, China, Japan and Taiwan. According to him, the relatively large roles of families and enterprises in meeting the welfare need in the region, which is regarded as central to the distinctiveness of its welfare model, is increasingly taken by the state, mainly because of structural changes such as rising unemployment rates, growing income inequalities, and rising populations of the retired generations.

However, many other authors note the limitations of applying elements and concepts of the traditional Western welfare state models to Japan or other emerging East Asian welfare states such as Korea and Taiwan. The critics warn against a "Western lens" (Hudson & Kühner, 2012, p. 35), "Swedocentric, Eurocentric and ethnocentric trends" (Takegawa, 2005, p. 160), and "a strong European bias" (Hudson & Hwang, 2014, p. 15). For example, Holliday (2000), in proposing a distinctive productivist welfare regime in addition to Esping-Andersen's three welfare types, suggests that the main features of the productivist world are 1) its social policy subordinate to economic policy, 2) its social rights remaining minimal, 3) its stratification effects reinforcing productivist elements and 4) state-market-family relationship premised on overriding economic growth objectives. Kim (2010) also suggests a distinct social welfare model in East Asia after identifying the region's 'surrogate social policy' measures such as producer support estimates for agricultural protection and mandatory private social spending for enterprise, after analysing the OECD data including South Korea and Japan. These measures "add up to make a difference between the East Asian countries and the other OECD members" (p. 1). In the end, "this coincidence of welfare expansion with economic liberalzation, a shift towards a more post-industrial economy and more pluralist polity, is highly significant and cannot easily fit within the classic 'Western' theories of welfare development" (Hudson and Hwang, 2013, p. 36).

Interestingly, Esping-Andersen, long reluctant to add another welfare types to his three archetypes for the sake of "the desired explanatory parsimony" (1999a, p. 88) states that East Asian welfare model can be interpreted two ways: as a hybrid of liberal and conservative or an emerging fourth welfare regime, in the preface to the Chinese version of *The Three Worlds of Welfare Capitalism* (Esping-Andersen, 1999b, p. 2.; cited by Lee & Ku, 2007, p. 199). Consequently, these two East Asian welfare states of South Korea and Japan, sometimes with other neighbouring states such as Taiwan or Singapore, are regarded as forming a distinct East Asian welfare regime in this thesis (e.g. Lee and Ku, 2007; Walker and Wong, 2005).

Turning to the second question on internal differences between Japanese and Korean welfare models, the former has been categorised as either a conservative welfare regime (e.g. Korpi & Palme, 1998) or a liberal regime (e.g. Scruggs & Allen, 2006). The majority of the studies within the 'welfare modelling business' (Abrahamson, 1999) include Japan with the other 17 Western welfare states (see Kim, 2015). This exceptional status of Japan, when compared with other East Asian nations, means the nation is often excluded from discussion on East Asia's laggard welfare regime (e.g. Jones, 1990; Wilding, 2008). Lee & Ku (2007) also place Japan between developmentalism and the Conservative welfare regime, while South Korea and Taiwan are seen as forming a new group apart from the other clusters of traditional welfare states, in their cluster analysis of 20 countries.

However, the majority of studies on the East Asian welfare model still include Japan (e.g. Aspalter, 2006; Kwon, 2009; Hudson and Hwang, 2013), as the Japanese model is seen as serving as an archetype for other East Asian late-coming welfare states. "For them, the Japanese model provided an obvious way forward, not only because of its demonstrable success but also because many key national institutions continued to be shaped by the intrusive colonial experience" (Holliday, 2005, p. 153). Another similar analysis suggests that "Japan's colonial legacies ran deep. Policy makers in Korea and Taiwan copied, first, Japan's medical insurance scheme, followed by the Japanese model for pensions and social care. Japan was and continues to be for Korea and Taiwan the most important source of social policy diffusion and learning." (Peng & Wong, 2008, p. 67)

Consequently, the South Korean and Japanese welfare models have been grouped together even within the region, named as follows: 'social insurance type countries' (Japan, Korea, Philippines, Taiwan and Thailand) among nine Asian nations (Park & Jung, 2009); 'development-unversalist' (Japan, Taiwan and Korea) among five Asian nations (Holliday, 2000); and an inclusive model (Korea, Taiwan, Japan and Thailand) among six Asian nations (Kwon, 2009). Given the historical and empirical analysis on the proximity of South Korean and Japanese welfare models, the two can be analysed within the East Asian welfare regime category in this thesis.

The next question is whether other nations such as Taiwan, another East Asian welfare state, arguably the closest model to Korean and Japanese welfare types, should also be included in this thesis. This is directly related to the third question, raised above, on the possible limitation of only the two nations forming an independent welfare regime. However, there are few comparable datasets for Taiwan as long as this thesis is concerned, because this thesis uses OECD databases and Taiwan is not a member of the OECD. On the other hand, it could be argued that the two cases of South Korea and Japan may fall short in the number of states, but not in terms of their combined population. The total population in the two East Asian welfare states (178 million) surpasses those of the Scandinavian welfare regime (27 million) and the South European welfare regime (132 million) (OECD, 2016b). It would be desirable to have an additional national case of Taiwan for the East Asian welfare regime in this thesis if the data was available, but the two nations of South Korea and Japan are arguably big enough at least in terms of size of population.

The Fourth and final question regarding whether inclusion of another East Asian welfare regime, such as Taiwan or Singapore, could change the findings of this thesis raises the issue of potentially selective case selection (see Kim, 2015). It can be argued that the thesis' findings on relatively good health outcomes in the East Asian welfare regime could be due to the choice of the two wealthiest nations of Japan and South Korea, not Taiwan or other East Asian nations. However, Karim et al. (2010), which include five East Asian territories of Japan, South Korea, Singapore, Taiwan and Hong Kong along with 25 other welfare states, find that East Asian territories overall show impressive aggregate health records. Japan has the longest life expectancy at 80.9 years with Singapore the second longest (80.4). Following Australia (80.1), Switzerland (80.0) and Sweden (80.0), Hong Kong (79.9) has the sixth longest life expectancy. It is difficult to sustain an argument of selection bias, as South Korea has the lowest life expectancy among the five East Asian cases. The inclusion of other developed East Asian cases would therefore *increase* East Asian aggregate health and the choice of Japan and South Korea as the representative of the East Asian welfare regime may underestimate otherwise better aggregate health outcomes.

1.4. Scope and Contribution

The primary focus is on the Scandinavian welfare regime and its performance with regard to aggregate health. This thesis also focuses on the East Asian welfare regime, which functions almost as a mirror image of the Scandinavian welfare regime. For example, Sweden and Japan, the most populous states from the respective welfare regimes, are deemed as “alike” in high life expectancy but “dramatically different” in other health determinants (Wilkinson, 1996, p. 213). As will be discussed in the following chapters, while the Scandinavian welfare regime turns out to underperform in aggregate health despite its relatively good health determinants of high GDP per capita, narrow disposable income inequality and clean environment, the East Asian welfare regime shows better-than-expected aggregate health despite its unfavourable health determinants such as relatively low GDP per capita and wide disposable income distribution. For the pair of counterintuitive outcomes, this thesis respectively suggests “the second Scandinavian puzzle’ and ‘the East Asian puzzle’. In the end, by identifying and discussing the two puzzles, this thesis reveals some limitation in current theories on aggregate health, and proposes some potential causes of the puzzles.

This thesis is designed to answer the research questions and to make substantive contributions to our understanding of the topic. The contributions include four methodological, five empirical and three conceptual ones.

Regarding the four methodological contributions, at first, a ‘review of reviews’ (Gough, Oliver, and Thomas, 2012) in Chapter 3 is conducted for the first time in cross-national health studies to review previous review articles in order to analyse previous literature on the largely controversial subjects. This method vividly illustrates the lack of a consensus on the Wilkinson Hypothesis and the Scandinavian performance in enhancing aggregate health.

Second, a ‘decomposition’ systematic review method is developed and introduced in Chapter 4 to capture and analyse multiple findings within each empirical journal article. This method is used to overcome potential limitation of previous systematic reviews that tend to oversimplify each article’s various findings. The decomposition method helps us to identify what this thesis proposes as ‘threshold effects’ over which the theoretically expected relationships between health determinants and aggregate health indicators reverse or vanish.

Third, this thesis develops a simple but theory-backed method to overcome arguably arbitrary case selection process. The method, elaborated in chapter 5, is applied in Chapters 6 and 7. The new method arguably settles the arguments over the questionable selection of a group of 'rich nations' and its logical consequences (see de Vogli, 2004; Babones, 2008; Pop et al., 2013).

Fourth, this thesis uses the OECD regional dataset covering all of the sub-national regions in OECD's 34 member states for the first time in the international comparative health study literature. The regional dataset can mitigate the chronic 'small-N' problem incurred due to the small number of welfare states (see Shalev, 2007; Esping-Andersen, 2007). There have been studies examining the regional variance in aggregate health but their foci were limited to single nations such as US states (e.g. Lynch et al., 1998) or Japanese prefectures (e.g. Shibuya, Hashimoto, and Yano, 2002). The large set of regional units enables this thesis to conduct a multiple regression with several independent variables.

The thesis has also five empirical contributions. First, despite the general expectation or belief that the Scandinavian welfare states have the best aggregate health levels, this thesis produces counterintuitive but consistent findings suggesting that they underperform in enhancing aggregate health when compared with other types of welfare states. The findings are consistent throughout Chapters 3, 4, 6 and 7. This thesis is arguably the first to acknowledge the relatively low level of Scandinavian aggregate health with valid evidence provided.

Second, another noticeable finding in this thesis is that the East Asian welfare regime (Japan and South Korea) has one of the best aggregate health when compared with other welfare regimes despite its relatively poor health determinants such as income and income inequalities. The East Asian welfare regime has the best health outcomes among the five welfare regimes compared in terms of female or old-age health indicators in

Chapters 6 and 7. This thesis is again the first case to acknowledge the outstanding East Asian health with valid evidence.

Third, this thesis also discovers that the relationships between income inequality, welfare regimes and aggregate health tend to reverse over a certain age threshold of the subject population. For example, in Chapter 4, the systematic review observes that the majority of the primary studies support the relationships when they use infant or child mortality as dependent variables but the results cannot be replicated for working-age or elderly mortality. The statistical outcomes in Chapter 6 and 7 also demonstrate that income inequality tends to have statistically significant negative associations more with infant mortality rate but less with old-age mortality rate. In addition, it is found that the hypothesised relationships sometimes show different patterns between female and male populations.

Fourth, regarding the role of economic growth, this thesis finds that GDP per capita is one of the most statistically significant determinants of aggregate health in Chapter 6 and 7. The findings runs counter to the Wilkinson Hypothesis that over a certain threshold of national income, a national economic growth ceases to influences its aggregate health (Wilkinson, 1992; Wilkinson and Pickett, 2009).

Fifth, this thesis tests the role of a long overlooked health determinant - market (pre-tax) income inequality - and finds that it has statistically significant associations with aggregate health. The majority of previous studies use disposable (post-tax) income inequality indicators because they measure the final amount of money pocketed by every household or individual. Few studies use market income inequality indicators (e.g. Sanmartin, 2003). However, the multiple regression in Chapter 7 also finds that market income Gini has as statistically significant associations with aggregate health as disposable income Gini. In Chapter 8, the pre-tax income distribution gives a clue for unpicking both the second Scandinavian puzzle and East Asian puzzle.

This thesis also makes the three following conceptual contributions. First, the thesis proposes a new concept of ‘the second Scandinavian puzzle’ after observing empirical findings on the relatively low level of Scandinavian aggregate health. The term is coined because Scandinavia’s puzzling underperformance is largely accepted in narrowing health inequalities within its population (i.e. the first Scandinavian puzzle). Altogether, the thesis suggests another term of ‘the dual Scandinavian puzzles’ for Scandinavia’s double underperformances in both aggregate health and health inequality.

Second, this thesis introduces the term ‘the East Asian puzzle’ after finding the better-than-expected health records in the East Asian welfare regime. Only a few studies include the East Asian welfare regime for cross-welfare regime comparative health study (e.g. Karim et al., 2010; Chuang et al., 2012), but they conclude that the East Asian health outcomes do “not have the worst” (Karim et al. 2010, p. 45) or “not have worse” (Chuang et al., 2012, p. e23) health when compared with other welfare regimes. However, the empirical findings in this thesis place the East Asian population as the healthiest people especially in terms of female or elderly health.

Third, this thesis also suggests a new term of ‘threshold effects’ involving age, gender, income and period. This thesis observes in Chapter 4, 6 and 7 that the each generational or gender health indicator often shows different patterns in response to an identical input of health determinants. The findings in the chapters demonstrate age-specific or gender-specific patterns that the hypothesized relationships between income inequality, welfare regimes and aggregate health often reverse between generations or genders. The hypothesized relationships also reverse over a certain threshold of GDP per capita or over a certain timing of observation. It is argued that without taking into account the four thresholds effects, we may oversimplify the dynamics of the relationships between income inequality, welfare regimes and aggregate health.

1.5. Structure and Content of Thesis

Chapter 2 provides an overview of the theoretical background on determinants of aggregate health as discussed in in the previous literature. This thesis finds in Chapter 3 the inconsistent conclusions on the relationship between income inequality and aggregate health (i.e. the Wilkinson Hypothesis) and also the contradictory conclusions on the relationship between welfare regimes and aggregate health (i.e. the second Scandinavian puzzle) after conducting the 'review of reviews'.

The subsequent systematic review, conducted in Chapter 4, demonstrates that the 48 empirical studies under the systematic review also do not present consistent evidence on the two research questions, confirming the lack of a consensus on the Wilkinson Hypothesis and corroborating the presence of the second Scandinavian puzzle. The new decomposition method also shows the four "threshold effects" involving age, gender, income and period.

Chapter 5 elaborates on and justifies the selection of methods, variables and cases for the following two chapters. The pooled time-series cross-section (TSCS) analysis in

Chapter 6 demonstrates; that the Scandinavian welfare regime does not show the best aggregate health outcomes despite its favourable health determinants, attesting to the second Scandinavian puzzle; that the East Asian welfare regime, paradoxically, shows the best aggregate health outcomes despite its mainly unfavourable health determinants, which corresponds with the East Asian puzzle; that GDP per capita and education are consistent and significant determinants of aggregate health; that other health determinants such as income inequality, alcohol consumption and public health spending are also statistically significant health determinants but their influences are limited to a certain generation or either male or female groups showing the age or gender threshold effects.

The multiple regression in Chapter 7 presents empirical evidence of 1) the second Scandinavian puzzle again, 2) the East Asian puzzle again, 3) GDP per capita as the consistent and significant health determinant again, 4) all income inequality-related indicators (disposasble income Gini, market income Gini or 'tax and transfer effects') emerging as the consistent and significant health determinants, and 5) other independent variables of air quality and unemployment rate having influences on only either one age group or gender group.

Then Chapter 8 discusses the key issues regarding the three empirical findings on the second Scandinavian puzzle, the East Asian puzzle and the Wilkinson Hypothesis. First of all, this chapter proposes two possible accounts on the second Scandinavian puzzle. The two accounts are the relatively wide market income inequality and the stagnant reduction in old-age mortality rates respectively. Second, this chapter also offers three potential accounts for the East Asian puzzle. The first and second accounts are the relatively narrow market income inequality rates and unemployment rates in East Asia.

The high intergenerational cohabitation is the third account that may explain the low old-age mortality rate in the region. Finally, regarding the Wilkinson Hypothesis, the chapter only partially supports the hypothesis because the findings in the previous chapters support the economic growth as a consistent health determinant but income inequality does not constantly have an effect on aggregate health indicators.

Finally, Chapter 9 summarises the main findings and argues that the thesis has twelve contributions: four methodological, five empirical and three conceptual ones. In addition, this thesis ends by discussing its five limitations and some academic and policy implications of the findings.

CHAPTER 2 THEORETICAL BACKGROUND

2-1. Introduction

In order to answer the research questions on the relationships between income inequality, welfare regimes and aggregate health, the first step is to trace the relevant theoretical background. In particular, it is necessary to ask why some researchers expect the Scandinavian welfare regime to record the best aggregate health among the welfare regime types. This chapter presents a systematic approach to reviewing various health determinants which have been suggested as accounting for regime difference in health. This inevitably involves an overview of the complex web of pathways proposed in the previous studies linking health determinants and aggregate health.

Previous theoretical accounts include, for example, the “artefact effect” (Gravelle, 1998), the “neighbourhood effects” (Lupton, 2003), the “collective effect” (Kawachi et al., 2002), the “place effect” (Macintyre, Ellaway, and Cummins, 2002), the “pollution effect and concavity effect” (Subramanian and Kawachi, 2004), the “area effect” (Dibben, Sigala, and Macfarlane, 2006), and the “contextual effect” (Pickett and Pearl, 2001; Frohlich et al., 2001). These sometimes appear to be different terms for a same or similar phenomenon, although perhaps stressing a new and notable aspect, and provide a fresh insight into relationships between health determinants and health outcomes. As this analysis of pathways is closely related to choice of variables and methods in this thesis, they should not be ignored or overlooked and need to be examined, selected and placed in a broader and clearer map of the theoretical landscape.

In this chapter, the challenging work of unravelling this theoretical 'tangle' starts by identifying and arranging eight proposed accounts regarding determinants of aggregate health. These are 1) artefact, 2) social selection, 3) materialistic, 4) behavioural/cultural, 5) psychosocial, 6) social capital, 7) environmental and 8) policy accounts. Notably, each account has its theoretical backing and supporting evidence, which will be discussed in the remainder of this chapter. As the second step of the 'untangling' process, we also identify three different levels through which the eight health determinants influence health: 1) individual, 2) regional and 3) national levels. Combined, a total of 24 (8 accounts x 3 levels) mathematically possible pathways emerge. As the next step, this researcher discusses all the 24 pathways individually and examine if we need to take into account or operationalise them in this thesis.

This chapter has four parts. The first part briefly introduces the eight theoretical accounts on health determinants together with the three levels, leading to the 24 possible pathways associated with different health outcomes. The second part reviews each of the 24 theoretical accounts individually to establish whether or not they can be operationalised. The third part proposes twelve variables based on the theoretical review and the fourth part is the brief conclusion of this chapter.

2-2. Unravelling the Theoretical Tangle

The Black Report (Department of Health and Social Security, 1980) suggests four causal pathways to account for social inequalities in health. They are (1) artefact, 2) social selection, 3) materialistic, and 4) behavioral/cultural accounts. The task of the report, which was commissioned by British government in 1977, was to shed light on the health inequalities among the different occupational groups in the United Kingdom. It identified the four hypothetical pathways to account for different health conditions in gradient social classes.

It needs to be noted that when health inequalities are related to variations of health in a society, aggregate health is related to its average health. In simple mathematics, in order to calculate a variation, we first need to establish an average. It logically follows then we cannot measure a society's health inequalities without establishing its aggregate health. Consequently, any determinant of health inequalities will also be a determinant of average health. For example, Gravelle (1998)'s theoretical model on the income-health relationship demonstrates that income inequality, one of the major health determinants, simultaneously affects aggregate health (average) as well as health inequality (variation). Then, the four accounts in Black Report can be applied to this thesis as all of them can be regarded as determinants of aggregate health.

Additional accounts are also needed because "Black Report's four models do not provide a satisfactory framework for explaining how health inequalities are produced" (Asthana and Halliday, 2006, p. 24). This chapter adds four more accounts to the Black Report's original accounts. They are 5) 'psychosocial' (Wilkinson, 1996, 1999), 6) 'social capital' (Kawachi, Kennedy, Lochner, and Prothrow-Stith, 1997; Rose, 2000), 7) 'environmental' (Rosen, 1993), and 8) 'policy' (Mackenbach, 2003; Subramanian and Kawachi, 2004) accounts.

In addition, the categories of various health determinants are suggested as layers of influence (Dahlgren and Whitehead, 1991; Whitehead and Dahlgren, 2001). For example, four layers are proposed:

"1) personal, behavioural factors,

2) individual interaction with peers and their immediate community,

3) the wider influences on a person's ability to maintain health in the third layer such as living and working conditions, food supplies, etc,

4) an overarching mediator of population health such as economic, cultural and environmental conditions prevailing in society as a whole" (Whitehead and Dahlgren, 2001, p. 313).

Inspired by the multi-layer model but modifying it in accordance with this thesis' research design, this chapter proposes another form of a 'vertical axis' depending on the following three scales of effects: individual, regional (sub-national) and national levels. The abovementioned eight health determinants can exert influence through the different levels. For example, a person's income is regarded as having an individual impact on the person, but air pollution impacts generally over a whole region. In another case, income inequality is not an individual but a collective indicator, which means that each person cannot have his or her own individual Gini indicator, mirroring its collective characteristics. Consequently, this chapter suggests discrete individual, regional and national levels through which the eight health determinants exert influence on health.

We can then combine the abovementioned eight accounts (artefact, social selection, materialist, behavioral/cultural, psychosocial, social capital, environment and policy) and the three tiers of individual, sub-national regional and national effects. This makes 24 mathematically possible pathways (3 tiers x 8 accounts) which are summarised in Table 2-1.

If we take one example of the fourth behavioural/cultural account as seen on the fifth column of the table, its impact can be conceptually and empirically divided into individual, regional and national health. At an individual level, a person's high tobacco consumption can result in bad health. In the regional level, a high rate of tobacco consumption in a region can be merely because of the high ratio of heavy smokers (i.e. compositional effect) but there might be other factors such as density of convenience stores in the region (Chuang et al., 2005). These physical environmental factors in a region may influence the "collective lifestyle" among people (Frohlich et al., 2001, p. 776), which leads to different smoking rates in different regions. In the broader national level, a national dietary culture, for example, the South Korean penchant for their nutritious dish, kimchi, can be a national-level behavioural or cultural health determinant. Likewise, the eight accounts may have distinctive multilevel pathways depending on their influence on individual, regional and national health.

The terminology warrants caution as the terms have different definitions and contents when used by different researchers in different levels. For example, each level has 'artefact' accounts but their meanings are different in all the three levels. To avoid the confusion, this researcher distinguishes the different versions by attaching I, II or III. However, if one account remains the same in its contents across different levels, the name would remain the same (for example, environmental account has the same Roman numeral of 'I' in both regional and national level) because its meaning remains the same in both levels.

Table 2-1. The Complete Set of Eight Health Determinant Accounts along Three Levels

	1. Artefact Account	2. Social Selection Account	3. Materialist Account	4. Behavioural /Cultural Account	5. Psycho-social Account	6. Social Capital Account	7. Environ-mental Account	8. Policy Account
Individual Level	**Artefact I**: little causal relationship between class and health (Black Report, 1980)	**Social Selection I**: the healthier a person, the richer one (Blane, Smith, & Bartley, 1993) i.e. health causes class	**Material I:** a person's socioeconomic position matters for health (Lynch, Smith, Kaplan, & House, 2000) i.e. class causes health	**B/C I**: "reckless or irresponsible behaviour… determinant of poor health status" (Black Report, 1980)	**Psychosocial I:** an individual's relative (often perceived) position on the social hierarchy matters (Wilkinson, 1999)	**Social Capital I:** An individual's social capital or social network matters (Rose, 2000)	Not applicable	Not applicable
Regional Level	**Artefact II:** Relationship between more equal income inequality and better health is only ecological fallacy (Gravelle, 1998)	**Social Selection II:** Some genetic characteristics of a group matter for health (Herrnstein and Murray, 1996)	**Material II:** Income inequality has independent contextual effects (Kawachi et al., 2004)	**B/C II** : Not just individual but 'collective lifestyle' matters for health (Frohlich et al., 2001)	**Psychosocial II:** 'Community stress' as "a state of ecological vulnerability" (Gee et al., 2004)	**Social Capital II:** Social trust or social capital influence health (Rostila, 2007)	**Environment I:** Local environment such as water or air has direct impacts on health (Rosen 1993)	**Policy Account I:** Regional health system has impact on health (Shi, 1992; Mackenbach, 2003)
National Level	**Artefact II, III**: Different national health measures lead to false comparison (Bloor, Samphier, & Prior, 1987)	Same as **Social Selection II**	Same as **Material II**	**B/C III:** Cultural differences account for cross-national health gap (Marmot & Smith, 1989)	No study found on 'national-level collective stress'	Same as **Social Capital II**	Same as **Environment I**	**Policy Account I, II:** health system (Macinko et al., 2004), welfare system (Subramanian & Kawachi, 2004) or welfare regime (Bambra, 2006a)

It should be also be cautioned that the web of relations between all the accounts all over the levels are quite porous, interactive and, consequently, controversial and there cannot be clear-cut dividing lines between them.

2-3. Examining the Pathways

This section now discusses the material in Table 2-1 in detail. This researcher presents the table first of all to provide a clear and organised overview of the ‘tangle’ of theoretical pathways proposed by previous studies as leading to different health outcomes. We here discuss each account contained in each cell of the table individually by introducing the claims of proponents of the particular account and the related criticism. The discussion is inevitably quite restricted and brief due to word limits.

2-3-1. Individual Pathways

These pathways affecting people’s health individually rather than collectively have long been the main focus of the traditional epidemiology (see Williams, 2003) which focuses on individual people’s behavioural and biological traits that result in personal disease or ill health (see Frohlich, Corin, and Potvin, 2001). As presented in the first row of Table 2-1, these individual pathways are divided into six accounts; artifact I, social selection I, behavioural/cultural I, materialist I, psychosocial I and social capital I. The other two environmental and policy accounts are omitted here because there may not be individual health determinants. For example, air pollution has an indiscriminate effect on people's health in a society. Consequently, its impact is collective rather than individual.

1) Artefact Account I

The Black Report hypothesizes that "both health and class are artificial variables thrown up by attempts to measure social phenomena and that the relationship between them may itself be an artefact of little causal significance" (Townsend and Davidson, 1982, p. 105). The social determinants such as income might be inferred to exert little, if any, influence on personal health. For example, Shaw et al. (1999, p. 89~90) point out that at the time of the publication of Black Report, the coding of individual person's social class at death was inaccurate due to technical reasons.

With such technical problems adjusted, subsequent studies find the artefact effect attenuates, rather than exaggerates, the difference in individual health (Smith, Blane, and Bartley, 1994). Since then, this artefact account I has rarely featured, for example, in the two British major reports on health inequalities (Acheson, 1998; Marmot et al., 2010). Subsequently it has begun to be "no longer considered as realistic" (Bambra, 2010, p. 399). This account will not be considered further in this thesis.

2) Social Selection

The idea here is that a person's health can influence his or her ability to move upward or downward in the social hierarchy: the healthier, the richer (Blane, Smith, and Bartley, 1993). In other words, "Occupational class is here relegated to the state of dependent variable and health acquires the greater degree of causal significance" (Townsend and Davidson, 1982, p.105). Chandola et al. (2003) distinguish two different approaches to prevent conceptual confusion of this account. The first is the health-related social mobility (presence of social selection) and the second the social selection hypothesis (the amount of contribution of social selection in explaining health inequalities). Then, with data from four phases of the Whitehall II study spanning 10 years, they conclude that social selection may be present but its contribution is not primary. In other words, health-related social mobility can be present, but its significance is minimal.

On the other hand, some claim that childhood health can play a role in the explanation of socio-economic differences in adulthood (Van De Mheen, Stronks, and Mackenbach, 1998; Stansfeld et al., 2011). These arguments are also related to the 'life course effect,' which refers to "how health status at any given age, for a given birth cohort, reflects not only contemporary conditions but embodiment of prior living circumstances, in utero onwards" (Kawachi, Subramanian, and Almeida-Filho, 2002, p. 650). In general, however, social selection factors are regarded as contributing relatively little or negligible to health inequalities (Lundberg, 1991, Blane, Smith, and Bartley, 1993; Marmot et al., 1997; Ki, 2009). This thesis therefore does not take into account this social selection pathway.

3) Materialistic Account I

The Black Report highlights the material/structural account as the foremost determinant of class difference in health. However, the account seems to have more than one component as the following sentences hint.

"Occupational class is multifaceted in 'advanced' societies, and apart from the variables most readily associated with socio-economic position - income, savings, property and housing - there are many other dimensions which can be expected to exert an active causal influence on health. People at work, for instance, encounter different material conditions and amenities, levels of danger and risk, degree of security and stability, association with other workers, levels of self-fulfillment and job satisfaction and physical and mental strain" (Townsend and Davidson, 1982, p. 109).

A careful reading of these sentences suggests that the account discusses two components of physical variables and 'other dimensions'. Macintyre (1997, p. 727) differentiates these as 'hard' ("Material, physical conditions") and 'soft' ("Physical and psychosocial features") accounts. The two versions each roughly correspond to two research groups: 'the neo-materialists' and what Macintyre, Ellaway, and Cummins (2002) call "the social cohesion/social capital theorists". The neo-materialists (Lynch, Smith, Kaplan, and House, 2000) emphasize structural causes of individual health, whereas the social capital theorists (Wilkinson, 1996; Marmot and Wilkinson, 2001) stress the psychosocial pathways from relative disadvantages to ill health.

This thesis also distinguishes between the two different explanations, naming them as materialistic (i.e. objective) and psychosocial (i.e. subjective) accounts, and here we will discuss the first individual-level materialistic account (then named here as Material Account I to distinguish it from other aggregate-level materialistic account). In this account, a person’s socioeconomic position matters, which is most frequently operationalized as “education, social class, occupation, income, housing characteristics, and wealth” (Regidor, 2006, p. 898).

One thorny issue regarding these indicators are that they are found to be generally not only associated with individual health but also interrelated to each other (Lahelma, Martikainen, Laaksonen, and Aittomäki, 2004). These interrelations form a web of interactive and mediating effects, posing a challenge to formation of a theoretical model when combining the variables. The combined effects of the multiple variables have rarely been explored, probably due to their complexity. Some studies claim that key materialistic variables such as ‘income, education and occupation’(e.g. Geyer, Hemström, Peter, and Vågerö, 2006) or ‘education, social class, income, status’ (e.g. Torssander and Erikson, 2010) have respectively independent pathways to health.

“Which of these yielded the strongest effects on health depended on type of health outcome in question. For diabetes, education was the strongest predictor and for all cause mortality it was income. Myocardial infarction morbidity and mortality showed a more mixed picture. In mutually adjusted analyses each social dimension had an independent effect on each health outcome” (Geyer, Hemström, Peter, and Vågerö, 2006, p. 804).

Given this caution, at least, the three most used indicators - income, occupation and education - need to be factored in as long as the relevant data are available.

In addition, with regard to income as a health determinant, this materialistic account I is strongly associated with the ‘absolute income hypothesis’ (Wagstaff and van Doorslaer,

2000) and ‘individual income interpretation’ (Lynch et al., 2000). The accounts deny any ‘contextual effects’ and only accept the ‘compositional effect’. According to the ‘compositional effect’ account, as long as two individuals have the same income, their health outcomes are assumed to be the same even if one lives in rich area and the other in relatively poor area. In other words, the claim is that when we control for individual socioeconomic status, there are no extra ‘area’ or ‘contextual' effects. This position is also closely associated with the regional and national level artefact II account to be discussed later.

4) Psychosocial accounts I

While the materialist account I is mainly associated with absolute individual conditions, this psychosocial account emphasises an individual’s (often perceived) relative position within the social hierarchy. Even the Black Report briefly takes note of this peculiar aspect of poverty by adding "poverty is also a relative concept” (Townsend and Davidson, 1982, p. 107). This individual psychosocial account encompasses three out of five of Wagstaff and van Doorslaer's (2000) hypotheses: the ‘relative income hypothesis (RIH)’, the ‘deprivation hypothesis (DH)’ and the ‘relative position hypothesis (RPH)’. As they put it, the difference between RIH and DH are “often unclear” (Wagstaff and van Doorslaer, 2000, p. 548) as DH is defined as “income relative to some poverty standard” (Lynch et al., 2004, p. 15). In other words, the three hypotheses are all related with this psychosocial account I because all three emphasize the relativity.

Wilkinson (1996, 1999) has particularly focused on individual's subjective perceptions of inequality leading to bad health outcomes: “Increasingly it looks like the most powerful influences on population health in the developed world are psychosocial” (Wilkinson, 1999, p. 492). According to Asthana and Halliday (2006, p. 26), “the psychosocial hypothesis became a conventional wisdom in the late 1990s”.

One technical difficulty regarding this account is in measuring the subjective perception because a person’s status in a social hierarchy and the person’s perception of it do not always match. For example, relative poverty rates (population below 50% of median income) were 9.1 percent in Sweden and 17.4 percent in the United States (US) in 2011 (OECD, 2014a), but the ratio of people who say they belong to the bottom 20% income group, paradoxically, was higher in Sweden (12.3%) than in US (8.3%) (World Value Survey Association, 2015). “The cultural meaning of economic inequality is also likely to vary and make a difference to outcomes” (Rowlingson, 2011, p. 26).

However, proponents of the psychosocial account such as Wilkinson and Pickett (2009, 2015) continue to use the objective income inequality indices rather than the subjective poverty indicators. This thesis tentatively follows their choice of ‘objective’ variables in Chapters 5 and 6 and discusses its possible limitation in Chapter 7.

5) Social Capital Account I

Putnam, Leonardi, and Nanetti (1994, p. 167) define social capital as “features of social organization, such as trust, norms, and networks, that can improve the efficiency of society by facilitating coordinated actions”. In addition to this collective definition of the concept, van der Gaag (2005, p. 2) emphasizes its individual aspect by stating the concept as “the collection of resources owned by the members of an individual’s personal social network, which may become available to the individual as a result of the history of these relationship”.

Rostila (2007) divides the social capital account affecting people’s health into two versions of ‘compositional effect (of social capital)’ and ‘contextual effect (of social capital)’. The former is summarized as “individual attributes and activities which contribute to social trust, which in turn might influence individual health” (p. 226). This narrow version will be discussed here as ‘Social Capital Account I’ while the broader one appear in the later Social Capital Account II part. On this social capital account I, positive relationships are found between individual social capital and their health (Rose, 2000; Poortinga, 2006; Kim, Baum, Ganz, Subramanian, and Kawachi, 2011).

However, it is unclear that social capital is an independent factor or merely a dependent factor of individual’s socioeconomic status. For example, one systematic review on sixty empirical studies finds “strong evidence to suggest that people with a lower socioeconomic status generally have lower levels of social capital” (Uphoff, Pickett, Cabieses, Small, and Wright, 2013). If there is a strong correlation found between individual socioeconomic status and social capital, we cannot choose both the variables. This thesis chooses the former as a significant variable because social capital accounts are found “not wholly adequate” (Cattell, 2001, p. 1501) or having “weaker associations with population health” (Muntaner et al., 2002, p. 619). This thesis does not take into account this social capital account I.

6) Behavioural/ Cultural Account I

The Black Report notes that "people harm themselves or their children by the excessive consumption of harmful commodities, refined foods, tobacco, and alcohol or by lack of exercise or by their under-utilization of preventive health care, vaccination, ante-natal surveillance or contraception" (Townsend and Davidson, 1982, p. 110). Macintyre (1997) again divides the account into hard and soft versions. The hard version emphasises the individual responsibility while the soft views the health damaging behaviours as collective culture. The hard version can be applied to this behavioural/cultural I account as individual behavioural account; "Health damaging behaviours freely chosen by individuals in different social classes explain away social class gradients (smoking, poor diet, inappropriate use of health services etc.)" (Macintyre, 1997, p. 727).

The Black Report also emphasises the individual responsibility for the risky life style such as "… unthinking, reckless or irresponsible behaviour or incautious life style as the moving determinant of poor health status" (Townsend and Davidson, 1982, p. 110). The individual and biological pathways from those behavioral factors to ill health can be clearer than those from any other health determinants. For example, smoking, "the foremost scourge of the twentieth century" (Ravenholt, 1990, p. 213) is estimated to incur more than 2.7 million deaths from lung cancer, more than 7 million deaths from cardiovascular disease, and more than 14 million deaths from all forms of diseases resulting from smoking (Ravenholt, 1990).

However, it is still unclear whether individual health-related behaviours could be treated as independent variables that have compositional effects on a national aggregate health, as the problematic individual behaviours are related to income gradients (Acheson, 1998; Shaw et al., 1999). In other words, they can be outcomes or mediating factors rather than independent determinants of ill health themselves (Barnett, Moon, and Kearns, 2004; Pampel, 2002). For example, Graham (1995) shows that smoking is a reasonable choice for people under heavy caring responsibilities and greater material disadvantage. "Stress may also affect health indirectly by leading to a more adverse profile of behaviours such as smoking and excess drinking" (Kawachi et al., 2002, p. 649). Wilkinson (1996, p. 2) asserts that "Nor does it seem as if the big health differences between societies can be explained by adding up individual behavioral risk factors such as smoking, exercise and diet".

7) Other Accounts at the Individual Level?

In addition to the six relevant accounts, some writers have suggested further accounts, which are difficult to classify in the framework above. For example, types or frequency of individual access to health care service may be another factor that can affect a person's health (Franks and Fiscella, 1998), but education is claimed to the predictor of the healthcare seeking behavior (Frie, Eikemo, and Knesebeck, 2010). At the regional and national level, however, the healthcare system may serve as independent variable affecting aggregate health. It will be discussed in later parts of this chapter.

Another pathway at the individual level can be termed the life course explanation that "combines aspects of the other explanations, thereby allowing different causal mechanisms and processes" (Bambra, 2011, p. 742). This can be a powerful account to predict individual difference in health (e.g. Smith, 2003), but has little contribution in suggesting independent variables in this thesis.

2-3-2. Regional (Sub-national) Pathways

Apart from individual pathways to ill health, researchers have focused on the independent impact of regional characteristics to the health of residents in collective ways. The hypothesized impact has various terms such as "neighbourhood effects" (Lupton, 2003), the "collective effect" (Kawachi et al., 2002) or the "place effect" (Macintyre et al., 2002), the "pollution effect and concavity effect" (Subramanian and Kawachi, 2004), the "area effect" (Dibben, Sigala, and Macfarlane, 2006) and most generally the "contextual effect" (Pickett and Pearl, 2001; Frohlich et al., 2001) and the "artefact effect" (Gravelle, 1998).

Conceptually, these accounts suggest that the social and physical environment will have an additional effect on health of the people residing within the territory, while a purely compositional account implies that similar types of people will have similar health conditions wherever they live (Asthana and Halliday, 2006). The list of regional approaches are similar in that they all attempt to distinguish between the compositional effect and possible additional effects, but often different in their unit of analysis, ranging from a small neighbourhood to a whole nation or even a group of nations.

The technical and theoretical problems in these accounts are that there are no clear dividing lines between relatively smaller regions (i.e. neighbourhood, villages or counties) and larger regions (i.e. countries). However, studies on relatively smaller regions often emphasize the physical living conditions of residents (Roberts, 1997; Chuang, Cubbin, Ahn, and Winkleby, 2005) while researchers on broader regions count different wide-scale variables such as policies (Starfield, Shi, and Macinko, 2005; Sjöberg, 2014) or social cohesion (Kim et al., 2011; Kennelly, O'Shea, and Garvey, 2003). This researcher focuses on sub-national geographical units in analysing these regional pathways.

1) Artefact Account II

As briefly discussed in the materialistic account I in the previous section, this artefact account endorses only the compositional effect and denies the contextual effect, labelling them only as a statistical 'artefact'. To avoid confusion, it should be noted that this artefact account II is different from artefact account I discussed earlier, which is mainly about erroneous coding. Artefact account II has its roots in Preston (1975)'s seminal work on the relationship between per capita national income and life expectancy demonstrating the health indicator began to progressively disassociate with the higher average income. Rodgers (1979) then demonstrates that the relationship between individual income and health status is nonlinear: the additional pound in income would enhance a person's health condition but only by a decreasing rate. One pound may be significant for the health of the poor, but serves little for a millionaire. This diminishing impact of additional one pound on health, in other words, the concave relationship between income increase and health enhancement, has critical implications when explaining this artefact effect on aggregate health.

It follows that the concave relationship leads to a logical expectation that the more money is given from the rich to the poor, the better the society's average health would be, because the poor person's gain is bigger than the rich person's loss in health. In other words, aggregate health would be better in a more equal society than in a less equal society with all other things assumed to be equal. Consequently, it is not the contextual effect but the underlying curvilinear function of the individual income-health relationship that accounts for the better health in a more equal society.

This is the reason Gravelle (1998, p. 382) goes on to contend that the association between equal society and better health may be "a statistical artefact resulting from the use of aggregate rather than individual data – an example of the "ecological fallacy"". The ecological fallacy is related to "the difference between ecological correlation and individual correlation" (Freedman, 1999, p. 1). For example, Robinson (1950) finds the statistically positive correlation between the percent of the population who are foreign-born and the percent of who are literate when the unit of analysis is 48 U.S. states. The counterintuitive correlation reverses and turns out to be negative at the individual level. The mismatch occurs because the foreign-born citizens tend to live in a state where the native citizens are more likely to be literate.

This artefact account, taking into account this ecological fallacy, has different names such as the 'absolute income hypothesis' (Wagstaff and van Doorslaer, 2000) or the 'individual income interpretation' (Lynch et al., 2000) but the point is the same in claiming that "there is no association between income inequality and health after proper control for absolute income at the individual level" (Lynch et al. 2004, p. 15). This hypothesis is also backed by some empirical findings (Fiscella and Franks, 1997; Muller, 2002). For example, Fiscella and Frank (1997), after conducting a longitudinal cohort study among 14,407 US citizens, state that community income inequality showed a significant association with subsequent community mortality, but vanished after adjusting for individual household income. There is a 'compositional effect' identified but not a 'collective effect'.

Although logically robust, the artefact accounts have faced criticisms. At first, even though the effect is described as artefact or even as "spurious" (Jen et al., 2009b, p. 643), the term 'artefact' is misleading because "it suggests that the potential for improving the health of the poor through income redistribution is a statistical illusion" (Subramanian and Kawachi, 2004, p. 80). For that reason, Subramanian and Kawachi (2004) suggest the new term of the 'concavity effect' instead of the 'artefact effect' to emphasize the actual impact of income inequality on people's health.

The second criticism is that the identification of this artefact effect does not necessarily imply rejection of the collective effect. In other words, there may be an independent collective effect operating 'in addition to' the artefact effect. To distinguish the collective effect from the artefact effect, Subramanian and Kawachi (2004, p. 80) proposes a new term of the "pollution effect". Some studies also identify the presence of the 'pollution effect' or the 'collective effect' even after adjusting for the individual income or education (e.g. Kennedy, Kawachi, Glass, and Prothrow-Stith, 1998; Wolfson et al., 1999). They contend that regional differences in health in United States cannot be substantially explained away as statistical artefact.

Despite all the controversy, the main point this thesis is interested in is that it is logically and empirically undeniable that income inequality is expected to be negatively associated with aggregate health, be it the outcome of 'artefact', 'pollution', 'context' or ‘concavity’. It follows that income inequality is an indispensable variable in accounting for aggregate health difference because even ‘artefact’ or ‘spurious’ association involves in actuality enhancement of health of relatively disadvantaged people (Subramanian and Kawachi, 2004a).

2) Social Selection (Genetic Difference) II

If one person’s health can contribute to their social position according to the Black Report’s natural/social selection account, the hypothesis might arguably be applied to a regional unit or possibly to an ethnic group. Then if members of a unit or a group share some genetic characteristics conducive to their good or bad physical condition (Herrnstein and Murray, 1996), they might be counted as another health determinant. In this context, two strands of studies have identified health differences between regions or ethnic groups.

The first is the gerontological or genetic approach to analysis of biological features of residents in particular regions especially well-known for their residents’ longevity, such as Japan’s Okinawa (Takata, Ishii, Suzuki, Sekiguchi, and Iri, 1987) or Italy’s Sardinia (Pes et al., 2004). The second is the aggregate health study approach to examining health differences between racial/ethnic groups. Deaton and Lubotsky (2003) argue that the correlation between income inequality and mortality rates across the cities and states of the US is confounded by the effects of racial composition. They claim that the ratio of the African-American population is positively correlated with the health indicator. Likewise, different health between different ethnic and racial groups have been reported in both the UK (Rudat, 1994; Nazroo, 1997) and US (Sorlie, Rogot, Anderson, Johnson, and Backlund, 1992).

However, few point a finger at the hereditary traits within and between the racial groups. Goodman (2000) contends that this larger level of social selection accounts requires acceptance of two disproved assumptions: “that genetic variation explains variation in disease and that genetic variation explains racial variation in disease... Using race as a proxy for genetic differences limits understandings of the complex interactions among political-economic processes, lived experiences, and human biologies” (p. 1699). Pearce, Foliaki, Sporle, and Cunningham (2004) also point out that the account is based on confusion between genetic, race and ethnicity and “genetic factors are important for health but are a small part of large and complex picture” (p. 1997). This thesis does not take into account this social selection account II.

3) Materialistic account II

This account argues that the aggregate health will be worse in more unequal communities than the relatively equal societies even after controlling for individual income. This account contradicts the abovementioned artefact account II (or ‘absolute income hypothesis’ (Wagstaff and van Doorslaer, 2000) or ‘artefact account’ (Gravelle, 1998)). Lynch and Kaplan (1997), after finding positive relationship between income inequality and mortality in the 50 US states after accounting for absolute levels of income, proposes a set of material, cultural, behavioural and psychosocial factors at the ecological level that may be linked with income inequality and influence health. Out of the four, their ‘ecological material factor’ is the one that is relevant for this part. (Despite their suggestion, they only discuss the material and psychosocial version and barely comment on the cultural and behavioural aspects.) They stress that “the higher inequality area also provides less equitable support for education, affordable housing, good roads or environmental protection” (p. 306).

Similarly, Kawachi and Kennedy (1999) propose three pathways linking income inequality and aggregate health. According to them, the first is “that income inequality is linked to disinvestment in human capital”; the second “that income inequality leads to the erosion of social capital”; the third “that income inequality leads directly to ill health via stressful social comparisons” (p. 220). Again, the first account is related to this ‘materialist account II’ and the second one for the next ‘social capital account II’, while the third one is on the abovementioned individual psychosocial account I.

Empirical studies confirm this materialist account II, especially when conducted for the US regions (Blakely, Kennedy, Glass, and Kawachi, 2000; Feng, Wang, Jones, and Li, 2012). In contrast, some studies repudiate any place-specific effects, especially for regions in other rich nations such as Denmark (Osler et al., 2002) and Japan (Shibuya, Hashimoto, and Yano, 2002). Whether or not there are any additional effects of income inequality present after controlling for individual absolute income, income inequality would be selected as a variable for the same reason as discussed in the artefact account II part.

4) Psychosocial Account II

It is not clear that the psychosocial stress incurred from relatively low income or social position has any collective effect on aggregate health even after controlling for individual psychosocial factors. That is to say, if there are two individual pathways (i.e. materialist I and psychosocial I accounts), there can be two collective pathways (i.e. materialist II and psychosocial II). However, studies clarifying the fourth ‘psychosocial II pathway’ could rarely be found. It is even questionable given the well-known controversy on the importance of individual psychosocial pathways between the neo-materialists (e.g. Lynch et al., 2000) and the researchers who Macintyre et al. (2002, p. 130) call “social cohesion/social capital theorists” such as Marmot and Wilkinson (2001). The latter social capital theorists seem to equate the collective psychosocial stress with the level of social capital.

However, as they define it, social capital is a concept on interpersonal relations between the social members, not the aggregate-level stress incurred from their material conditions. For example, according to them, “social capital describes the links between individuals: links that bind and connect people within and between communities” (Marmot et al., 2010, p. 30) or “those features of social organization-such as the extent of interpersonal trust between citizens, norms of reciprocity, and vibrancy of civic associations-that facilitate cooperation for mutual benefit” (Kawachi and Kennedy, 1999, p. 221). It can be inferred that the social capital accounts do not take into account this possibly ‘aggregate level psychosocial stress’ or ‘collective stress’ which at least needs to be tested to see its presence or absence.

Some empirical studies, albeit without clear conceptual acknowledgment of this possible account, give clues. Subramanian and Kawachi (2006), based on pooled data from the 1995 and 1997 US population surveys, claim that income inequality can be more harmful for the health of more advantaged groups such as whites and individuals with incomes greater than $75,000 when controlling for the composition effect of individual demographic traits. “This would suggest some sort of ‘‘social pollution’’ effect of income inequality that appears to affect every group exposed in a similar manner” (p. 149). As this “some sort of “social pollution” effect” has logically little to do with any of materialist account I, II, psychosocial account I, or the following social capital account I, II, this independent effect from the income inequality might be related to this psychosocial account II. (Social capital account II may be related to this “some sort of “social pollution” effect”, but could not explain why relatively advantaged people are more affected by income inequality.)

Gee et al. (2004) propose a concept of ‘community stress’ as “a state of ecological vulnerability” but their concept is related to environmental pollutants, structural process, community stressors, and neighbourhood resources, not the relative socioeconomic status of the community. This under-discussed effect seems to be a gap among thousands of related health studies and needs more analyses in the future. This thesis may not be able to operationalize this concept until its presence is studied further.

5) Social Capital Account II

As seen in the previous materialist account II, out of the Kawachi and Kennedy’s (1999) three pathways directly linking income inequality and aggregate health, the second “that income inequality leads to the erosion of social capital” (p. 220) is related to this account. Rostila (2007) also proposes ‘contextual effect (of social capital)’ that “social trust could also influence the political and social environment in a society, area or welfare regime and, as a consequence, influence health indirectly” (p. 227).

This somewhat abstract concept might have a measurement issue. For example, Kawachi, Kennedy, Lochner, and Prothrow-Stith (1997) measure it based on degree of mistrust or membership of voluntary associations published in a survey. In similar ways of measurement, some empirical research identify, in accounting for regional difference in health, ‘social capital’ (Subramanian, Kawachi, and Kennedy, 2001) in the US states or ‘neighbourhood cohesion’ (Ellaway, Macintyre, and Kearns, 2001) in communities in Glasgow. Focusing on thirty districts in Saskatchewan province in Canada, Veenstra's (2002) finding is more subtle that “Income inequality was not as strongly related to age-standardised mortality after controlling for social capital, and vice versa, suggesting the two may be co-mingled somehow when it comes to population health, although they were not significantly related to one another” (p. 849).

The thesis chooses not operationalise this rather elusive concept as a variable because, as Rostila (2007, p. 226) comments, “the mechanisms linking contextual trust and health are still vague and unexplored”.

6) Behavioural/ Cultural Account II

Collective behavioural/cultural pathways can at least partly account for regional health variation. “Places with high levels of smoking, for example, may simply be composed of more people with individual characteristics indicating a predisposition to smoking. Alternatively, all people in that place, regardless of their individual, personal characteristics may be affected by contextual, ecological factors (e.g. a regional culture that encourages smoking)" (Duncan, Jones, and Moon, 1993, p. 727). The regional smoking-encouraging culture is also not in isolation from other factors as it can be related to, again, physical surroundings such as density of convenience store in the region (Chuang et al., 2005) or regional affluence (Ross, 2000).

Given this complex web of mutual influence, Frohlich et al., (2001) propose a concept of 'collective lifestyle' defined as "an expression of a shared way of relating and acting in a given environment" (p. 776). The concept focuses on the tripartite "relationship between agency (the ability for people to deploy a range of causal powers), practices (the activities that make and transform the world we live in) and social structure (the rules and resources in society)" (p. 776). Then the collective lifestyle can involve health-related behaviours such as smoking, drinking, eating and exercise. This thesis uses any data on the collective lifestyles as long as there are available and reliable datasets.

7) Environmental Account

This account can emerge arguably only from the regional or national level as it collectively influences people as a whole. In 19^{th} century London, for example, the number of deaths from cholera was found to be associated with water pollution in different parts of the Thames River (Rosen, 1993). The quality of people's local environment, mainly water and air, has a direct impact on their collective health (OECD, 2015a). "An unspoiled environment is a source of satisfaction, improves mental well-being, allows people to recover from the stress of everyday life and to perform physical activity" (OECD, 2015a). The pathways are not individual but collective. Illustrating the five main local features that might be promoting or damaging local health, Macintyre et al. (2002) count as the first factor 'Physical features of the environment shared by all residents in a locality' which "include the quality of air and water, latitude, climate, etc. and are likely to be shared by neighbourhoods across a wide area" (p.131). The quality of air and water is probably one of the least undisputable independent factors that can directly influence local aggregate health, but surprisingly few empirical studies test its influence on aggregate health.

8) Policy Account I, II

Both health policy and welfare institutions can be determinants of aggregate health at the regional level. For example, some researchers (Shi, 1992; Mackenbach, 2003) claim that regional level health systems can account for differences in health between regions in United States and the Netherlands respectively. However, as these determinants most commonly operate at the national level in developed nations, they are dealt with in detail later in this thesis (see sections 2-3-3-8, 2-3-3-9 below).

2-3-3. National Pathways

These macro pathways linking the wider-level health determinants to the national aggregate health have similarities with those of the regional pathways with only some additional characteristics attached to, for example, artefact, materialist and behavioural/cultural accounts in this national level. The last policy pathway may need particular review as the difference in policies are made clear especially in cross-national comparison and they are contended to have significant influence each nation's aggregate health (Macinko, Shi, and Starfield, 2004). Again, it can be seen as controversial to divide the level of analysis into the sub-national regions and national entities, given that the wide variety of regions or nations in size. For example, the tiny state of Luxembourg with only half million population is much smaller than the US state of California with more than 30 million residents. Given these technical complications, the distinction between sub-national region and nation is basically conceptual. As this thesis consists of comparative cross-national and cross-regional analyses, we need to examine the sub-national regional level and then turn to the broader national level.

1) Artefact Account II, III

The artefact II account is applicable to the national level as well. If the curvilinear relationship between individual income and health would make every increase in income lead to a gradually decreasing increase in health, this will also result in the 'artefact' relationship between income inequality and aggregate health not only at the regional level but also at the national level (Gravelle, 1998; Jen et al., 2009a).

In addition, some commentators also suggest another artefact account (Huijts and Eikemo, 2009; Bambra, 2011), which can be termed here as artefact account III. In analysing the cross-national difference in health outcomes, they suggest that some factors may result in misleading outcomes such as the differences in “the data used and methods of measurement” (Bambra, 2011, p. 742) between nations. Combined, the artefact account III can indicate the possible misleading comparison between national health indicators due to different datasets, different selection of independent and health indicators as dependent variables, different selection of case nations, difference in people’s subjective evaluation of their own health. These problems can happen to the sub-regional level but will probably be more apparent and influential in this wider cross-national comparison. In addition, this artefact account III have some similarities with the artefact account I as the latter is related to "Different recording or coding conventions between... datasets” (Bloor, Samphier, and Prior, 1987). This artefact account III implies the need for caution in the selection of dataset, variables and methods to be discussed in the next chapter.

2) Social Selection (Genetic Difference) II

There is little reason to hypothesise that the genetic account can provide clues for differences in national aggregate health when it already fails in the regional level. As we see in another related study, Marmot and Smith (1989) provide a noticeable example where the ethnic Japanese living in Japan, Honolulu and San Francisco Bay area turn out to have quite different prevalence of coronary heart disease at 2.5%, 3.5% and 4.5%

respectively. “A genetic explanation would not account for the dramatic improvement in life expectancy over the past 20 years. It also would not account for the changes in mortality patterns that have occurred among Japanese migrants to the United States of America” (Marmot and Smith, 1989, p. 1549). The findings indicate that not the genetic differences but the other factors such as environmental or socioeconomic factors, discussed in the chapter, are the meaningful determinants of aggregate health. This thesis does not operationalise this account.

3) Materialistic Account II

The materialist account II which was discussed at the regional level can be applied to the national aggregate health.

4) Psychosocial Account II

As discussed in the regional psychosocial account, it is not clear if there is any conceptual division between ‘any psychosocial impact on the whole population over individuals’ (psychosocial account II) and social capital account in previous literature. The concepts of psychosocial impact and social capital have been confusingly mixed (e.g. Kavanagh, Turrell, and Subramanian, 2006) despite their clear difference as discussed in the previous section on the psychosocial account II. One technical issue is how to measure these psychosocial factors, if there is any, particularly in cross-national studies. For example, Runciman (1966) suggests that in forming the notion of their own relative deprivation, people tend to compare themselves with their peers rather than people as a whole. Citing this study, Rowlingson (2011) points that “so the broader income distribution may not be particularly relevant” (p. 23) as the aggregate psychosocial indicator.

Goldthorpe (2009) also finds that the concepts of status are different between nations. For example, Japan, one of prototype nations with narrow income inequality and consequently presumed to incur less status-related stress to its people, turns out to have “marked status hierarchy… one that is to an unusual degree formalized” (p. 8). Therefore, “the inferences that are made from the available data on income distributions to inequalities of status and their consequences are often of a doubtful kind” (p. 1). Given that the relationships between income inequality and aggregate psychosocial stress in this cross-national level may be quite open to controversy yet, it is another thorny issue of how to gauge the national-level psychosocial stress. Consequently, this thesis could not operationalize this important but complex concept.

5) Social Capital Account II

Studies linking the contextual social capital and aggregate health in the cross-national level studies show contrasting outcomes with the associations contended to be “not related” (Poortinga, 2006), “inconsistent” (Mansyur, Amick, Harrist, and Franzini, 2008), “strongly linked” (Helliwell and Putnam, 2004) and “strong” (Rostila, 2007). For example, Poortinga (2006) analyses 21 European countries in the European Social

Survey and finds that national-level social trust is not associated with individual self-rated health, after adjusting for compositional differences, whereas the effects of individual's social trust affects her or his health. On the other hand, Helliwell and Putnam (2004) contend, based on data from 49 nations, that social capital supports both physical health and subjective well–being of the respondents. Overall, regarding all the three-level pathways linking social capital and health, we can learn that 1) correlations between socioeconomic status and social capital is identified in individual level, making it hard to be treated as an independent variable, 2) in regional and national levels, it has limitations of being "vague and unexplored" (Rostila, 2007, p. 226) and 3) related findings are contrasting in the last cross-national level. Based on the three limitations, this thesis decides not to operationalize the accounts.

6) Behavioural/ Cultural Account III

At the cross-national level, the cultural differences emerge as one factor accounting for difference in aggregate health between nations. In examining the factors behind the Japanese longevity, Marmot and Smith (1989) note that diet is probably an important factor in lower rate of coronary disease and colon and breast cancer in Japan. Between nations, the class gradient in vegetable and fruits consumption is also found to be different. In a review of empirical studies on European dietary culture, the relationship between high vegetable consumption and high educational level is observed in the northern and western European nations but the relationships reversed in some southern and eastern European nations (Roos, Johansson, Kasmel, Klumbiené, and Prättälä, 2001; Prättälä et al., 2009). "Cultural characteristics and ingrained behaviours undoubtedly play a role also, at least in particular countries and should be included where relevant" (Starfield, 2007, p. 1360). However, few studies include the cultural/behavioural factors in the cross-national comparative health studies with only some exceptions (e.g. Armstrong and Doll, 1975; Stanistreet, Bambra, and Scott-Samuel, 2005). This thesis operationalises this account as long as relevant dataset is available.

7) Environmental Account

Hertz, Hebert, and Landon (1994) conduct a comparative study on 66 nations with three dependent variables (infant & maternal mortality and life expectancy) and a range of independent variables (medical resource availability, GNP per capita, literacy rates, growth in the labour force and provision of sanitation facilities and safe water). Their explanatory stepwise regression models show that the provision of sanitation facilities has the strongest association with all the three dependent variables. Water quality was also one of three independent variables strongly associated with all the health indicators. Despite not clearly presenting its criteria in selecting the 66 nations, this study include some low-income nations, which may be the reason the environmental variables account for the majority of variations in health between nations. However, given the variations in water quality and air pollution even among the developed nations (OECD, 2015a) and this related data availability, it is surprisingly difficult to find studies on the relationships between environmental factors and nation-level aggregate health.

8) Policy Account I – Health System

Social epidemiologists have been sceptical on the role of medical system in boosting aggregate health. They view the system either as having an adverse effect on people's health by creating rather than curing illness or disease (Illich, 1976) or as serving relatively little at best when compared with other social and structural health determinants, hence figuratively called "the ambulance waiting at the bottom of the cliff" (Daniels, 2007, p. 76). McKeown (1979) also note that the biggest improvement in developed societies is not due to medical system but due to better sanitation and food. The views are consistent with the claim that "Neither medical care nor genetics explains why one country is healthier than another, or why most countries gain two or three years of life expectancy with each decade that passes" (Wilkinson, 1996, p. 2).

However, the general skepticism especially in the 1960s or 1970s is because the past health care still had relatively little to offer and the recent development in medical system seem to make difference (Nolte and McKee, 2004). Empirical studies have shown the health system's contribution to aggregate health in developed nations in both regional level (Shi, 1992; Mackenbach, 2003; Shi, Macinko, Starfield, Politzer, and Xu, 2005) and national level (Macinko et al., 2004; Chung and Muntaner, 2008). For example, based on fixed-effects multivariate regression on 19 OECD nations, Macinko et al. (2004) report that health system variables, including the methods of healthcare financing and the supply of physicians, significantly attenuate the effects of wage inequality on infant mortality. Chung and Muntaner (2008) contend that total public medical coverage was the most significant predictor of their mortality indicators among other factors such as income inequality after analyzing time-series data from 19 OECD nations. The health system in the national level may need, if possible, to be tested on its influence on aggregate health.

9) Policy Account II – Welfare Policies

Subramanian and Kawachi (2004) propose a "policy pathway" in addition to their 'structural pathway' and 'social cohesion pathway' all of which link income inequality and health. According to them, the first pathway indicates, "the adverse influence of income inequality may operate through formulation and implementation of general social policies, as well as through health related policies" (p. 87). They take such possible mediating variables as "primary health care indicators, welfare spending, child care, food assistance, vocational training, remedial training, health insurance, early childhood education, disability assistance, tax policy" (p. 87). Including the health-related systems, they might arguably be broadly termed as welfare policies or welfare generosity. Empirical studies also claim some of the policies can be determinants of aggregate health, including occupational policy (Lipscomb, Loomis, McDonald, Argue, and Wing, 2006), pension policy (Lundberg et al., 2008) and family policy (Engster and Stensoeta, 2011; Ferrarini and Norström, 2010). However, still "there have been relatively few studies which have tested the impact of welfare provision" (Rowlingson, 2011, p. 27). The issue regarding how to combine and conceptualise the various welfare policy and related delivery in either cash or services is another challenge in identifying these policy pathways. This issue of operationalizing these pathways is discussed in section 5-2-1.

10) Policy Account III – the Welfare Regime Account

“The field of (macro) social epidemiology suffers from lack of comprehensive models” (Chung and Muntaner, 2008, p. 17, quoting Macinko, Shi, Starfield, and Wulu, 2003). Then “social epidemiologists have increasingly started to look to the comparative social policy literature to help construct explanations of the differences in health that exist between countries” (Karim, Eikemo, and Bambra, 2010, p. 45). The concept of welfare regimes (Esping-Andersen, 1990, 1999a) has emerged as a powerful concept to ‘go beyond’ the previous simple pathway account to cover a broader set of the social determinants of aggregate health (e.g. Coburn, 2004; Bambra, 2006a). Even though the concept has been under criticism for its methodological shortcomings (Bambra, 2006b), unclear conceptual basis (Powell, 2015) and ignorance of service delivery especially health service (Jensen, 2008), researchers began to use welfare regime to account for aggregate health differences between nations.

Bambra (2006a) is arguably the first to analyse the relationship between in welfare regime characteristics and aggregate health. She contends that the statistically significant negative association is found between labour market decommodification (a key indicator of welfare regime) and infant mortality rates (a proxy of aggregate health) with data of 18 rich nations after hypothesising that the former indicator would mediate the negative impact of income inequality on aggregate health. The decommodification index was calculated based on each nation's three income maintenance programs: pensions, unemployment and sickness. Coburn's (2004) model regards welfare regime as an independent factor rather than a mediating factor by stating, "income inequality is a consequence, not the determinant, of societal 'types'"(p. 43). Then he claims that the Social Democratic regime has less absolute and relative poverty which in turn is related to health differences within its residents especially when compared with the Liberal regime.

2-4. Operationalisation of Accounts

Given most of these accounts briefly examined individually, the next step would be to check if the accounts can be operationalised and, if so, how to do it for this thesis in this cross-national or cross-regional research design.

To summarize the brief review above in view of applying the accounts practically, **the artefact account I and III** is an important caution to minimize any statistical 'noise' that can affect research outcomes. However, they can hardly be operationalised in any respect for this thesis. On the other hand, according to **the artefact account II**, narrow income inequalities in a society should contribute to the better aggregate health. Then a society's income inequality indicator such as ① Gini index (e.g. Kondo et al., 2009; Ploubidis, Dale, and Grundy, 2012) should be examined in relation to aggregate health. All of **social selection accounts I and II** in any level is refuted (Marmot et al., 1997; Marmot and Smith, 1989). **The materialist account I** may, as discussed, require at least the three most used indicators, ② income (e.g. Pop, Ingen, and Oorschot, 2013; Torre and Myrskylä, 2014), ③ education (e.g. Doorslaer and Koolman, 2004) and ④ occupation (e.g. Mackenbach et al., 2008; Richter et al., 2012) as they have both independent and interactive influence an individual's health (Geyer, Hemström, Peter, and Vågerö, 2006). **The materialistic account II** supports the presence of 'contextual effect' or 'pollution effect' in addition to the compositional effect of income inequality. To take into account the additional effect, the inclusion of ① Gini index as an independent variable is necessary again. As long as individual income or income inequality cannot be a valid indicator for either **the psychosocial account I or II** (Goldthorpe, 2010), we may have to wait for any relevant data or income inequality may serve as a proxy variable for the psychosocial impacts.

For **the social capital account I**, this thesis follows the contention the individual-level social capital may not be an independent variable as long as it is related to the individual socioeconomic status (Uphoff, Pickett, Cabieses, Small, and Wright, 2013). **The aggregate social capital account II**, despite some empirical evidences on its relationship with aggregate health, still seems to be "vague and unexplored" (Rostila, 2007, p. 226) for its pathway to the aggregate health to be operationalised in this thesis.

The behavioural/cultural version I may not be operationalised as they are contended to be rather dependent variables than independent variables (Barnett, Moon, and Kearns, 2004; Pampel, 2002). However, **the behavioural/cultural accounts II and III,** taking into consideration between-region or cross-national differences, could be arguably one of the strongest health determinants that can be operationalised as ⑤ alcohol consumption (e.g. Hoffmeister, Schelp, Mensink, Dietz, and Böhning, 1999) ⑥ smoking population rate (e.g. Lawrence, Mitrou, and Zubrick, 2009) and ⑦ dietary characteristics (e.g. Armstrong and Doll, 1975).

Environmental factors such as ⑧ water quality (Benova, Cumming, and Campbell, 2014) and ⑨ air quality (O'Neill et al., 2003) can be used as variables relevant for **the environmental account**. In addition, **the three policy accounts** show strong theoretical and empirical evidence that call for use of such variables as ⑩ public health expenditure (e.g. Regidor et al., 2011), ⑪'redistributive effect from taxes and transfers (Luebker, 2012; Joumard, Pisu, and Bloch, 2012) and ⑫ decommodification index (e.g. Coburn, 2004; Bambra, 2006b). Empirical studies tested several welfare policies such as occupational policy (Lipscomb et al., 2006), pension policy (Lundberg et al., 2008) and family policy (Engster and Stensoeta, 2011; Ferrarini and Norström, 2010), but we cannot use every individual policy as distinct independent variable. ⑪′redistributive effect from taxes and transfers′ which is the difference deducted from market income Gini coefficient by disposable income Gini coefficient (Luebker, 2012; Joumard, Pisu, and Bloch, 2012) may represent the combined re-distributional effects of one society's welfare policies.

These decisions and selections on the variables and related theories are summarised in Table 2-2, which builds on the framework presented earlier in Table 2-1. Table 2-1 demonstrates the 24 pathways suggested by the previous studies, while Table 2-2 presents how the 24 pathways can (or cannot) be operationalised. (Appendix 2-1 contains further details) In Chapter 5, further discussion will be given on the selection and justification of the health determinant variables.

2-5. Conclusion

This theoretical review chapter takes the daunting challenge of unravelling the 'tangle' of pathways linking various health determinants and individual, regional and national health. As the first step, this chapter combines the eight accounts, suggested by the previous literature: 1) artefact, 2) social selection, 3) materialistic, 4) behavioural/cultural, 5) psychosocial, 6) social capital, 7) environmental and 8) policy accounts. The second step of the 'untangling' process is to identify the three different levels through which the eight health determinants work: 1) individual, 2) regional and 3) national levels. Combined, a total of 24 (= 8 accounts x 3 levels) mathematically possible pathways are discussed.

Table 2-2. The Selection of Variables from Table 2-1

	1. Artefact Account	2. Social Selection Account	3. Materialist Account	4. Behavioural /Cultural Account	5. Psycho-social Account	6. Social Capital Account	7. Environ-mental Account	8. Policy Account
Individual Level	**Artefact I**: "No longer considered as realistic" (Bambra, 2010, p. 399). => NONE	**Social Selection I**: Negligible contribution (Marmot et al., 1997; Ki, 2009) => NONE	**Material I:** Socioeconomic status matters (Geyer et al. 2006) => ②Income, ③ Education, ④ Occupation	**B/C I**: Maybe not independent determinants of ill health (Pampel, 2002) => NONE	**Psychosocial I:** Income inequality index used (Wilkinson & Pickett, 2009) => ① Gini index, ② Income	**Social Capital I:** "not wholly adequate" (Cattell, 2001, p. 1501) => NONE	Maybe not individual health determinant => NONE	Maybe not individual health determinant => NONE
Regional Level	**Artefact II:** Income inequality effect in not artefactual, but actual (Subramanian & Kawachi, 2004) => ① Gini index	**Social Selection II:** "the account is based on confusion" (Pearce et al., 2004, p. 1997) => NONE	**Material II:** Income inequality has contextual effects (Kawachi et al., 2004) => ① Gini index	**B/C II** : 'collective lifestyle' matters (Frohlich et al., 2001) => ⑤ alcohol, ⑥ smoking, ⑦ diet	**Psychosocial II:** Few studies use this concept of collective stress => NONE	**Social Capital II:** Effect "vague and unexplored" (Rostila, 2007, p. 226) => NONE	**Environment I:** Water or air has direct impacts (OECD, 2015a) => ⑧ water, ⑨ air	**Poicy Account I:** Regional health system has impact on health (Shi, 1992; Mackenbach, 2003) => ⑩ medical system
National Level	**Artefact II, III**: Income inequality has independent contextual effect (Subramanian & Kawachi, 2004) => ① Gini index	Same with **Social Selection II** => NONE	Same with **Material II** => ① Gini index	**B/C III:** Cultural differences matter (Marmot & Smith,1989) => ⑤ alcohol, ⑥ smoking, ⑦ diet	Few study found on 'national-level collective stress' => NONE	Same with **Social Capital II** => NONE	Same with **Environment I** => ⑧ water, ⑨ air	**Policy Account I, II:** Health system (Macinko et al., 2004), welfare system (Subramanian & Kawachi, 2004) welfare regime (Bambra, 2006a) matter => ⑩ medical system, ⑪ welfare policy, ⑫ welfare regime

The third step is to select the health determinants to be operationalized in this thesis based on its theoretical robustness and related data availability. In the end, 12 health determinants can be set up in this thesis as potential independent variables. They are ① income inequality, ② income, ③ education, ④ occupation , ⑤ alcohol consumption, ⑥ smoking, ⑦ dietary characteristics, ⑧ water quality, ⑨ air quality, ⑩ public health expenditure, ⑪'redistributive effect from taxes and transfers' and ⑫ decommodification index.

With this comprehensive mapping of complex pathways, this chapter provides theoretical backgrounds to the ensuing Chapters 3 and 4 where this researcher reviews how a significant number of previous empirical findings contradict the theoretical expectations, which this chapter discusses. Taken together, these two chapters provide guidelines for selection of variables, case and methods in this thesis.

CHAPTER 3 REVIEW OF REVIEWS

3-1. Introduction

Scandinavian nations have the lowest income inequalities and the most generous welfare system among wealthy nations (Esping-Andersen, 1999a; Gottschalk and Smeeding, 2000). The group of nations, termed as Social Democratic welfare regime (Esping-Andersen, 1990), show the best outcomes in terms of promoting efficiency, reducing poverty and promoting equality when compared with the other welfare regimes (Goodin, Headey, Muffels, and Dirven, 1999). If we narrow down the focus to health outcomes, the Scandinavian states are especially expected to show the best aggregate health outcomes and, at the same time, the lowest health inequalities compared with other wealthy welfare states due to the egalitarian labour market and strong redistribution policies (e.g. Bambra, 2011; Hurrelmann, Rathmann, and Richter, 2010). In other words, the Scandinavian welfare regime is expected to 'kill the two birds with one stone' as Gravelle (1998)'s famous theoretical model illustrates that narrow income inequality would logically lead to narrower health inequality (variation) and better aggregate health (average).

However, empirical findings often repudiate the hypotheses for both health inequalities and aggregate health. First of all, on health inequalities, Scandinavian nations do not show the narrowest health inequalities according to the majority of literature (see Brennenstuhl, Quesnel-Vallée, and McDonough, 2012), or the expected effect is inconsistent at best (Mackenbach, 2012). Second, regarding aggregate health, empirical findings also suggest Scandinavia's unexpectedly low level of aggregate health as well (Rostila, 2007; Kangas, 2010). Even two systematic reviews on the relationships (Brennenstuhl, Quesnel-Vallée, and McDonough, 2012, Bergqvist, Yngwe, and Lundberg, 2013) do not confirm the Scandinavian regime's good health. It is dubious that Scandinavian welfare regime excels in either of health inequalities or aggregate health.

Given this possible 'dual' underperformance of Scandinavian states, it is noticeable and surprising that only the first underperformance (health inequalities) has received wider and more intense attention while the second underperformance (aggregate health) is questionably overlooked or ignored. The first on health inequalities has long been under the spotlight in international comparative health studies and the counterintuitive findings are called as Scandinavian "puzzling finding" (Lahelma and Lundberg, 2009, p. 445), 'puzzle' (Bambra, 2011), 'paradox' (Hurrelmann, Rathmann, and Richter, 2010) and even regarded as 'the greatest disappointment' (Mackenbach, 2012). Commentators also interrogate the puzzle and attempt to find clues to possible answers (e.g. Huijts and Eikemo, 2009; Popham, Dibben, and Bambra, 2013).

On the other hand, Scandinavia's relatively low aggregate health has strangely little attention compared with its disappointing records in health inequalities. Consequently, there is still the general assumption prevailing that the Scandinavian welfare regime would be the best in terms of aggregate health. That is why many comparative health researchers take for granted the Scandinavian excellence for example, by stating that "population health is enhanced by the relatively generous and universal welfare provision of the Scandinavian countries" (Bambra, 2011, p. 2) or "the universalistic and redistributive approach taken in the Nordic countries makes positive overall health outcomes” (Huijts and Eikemo, 2009, p. 2).

The focus of this chapter is the discrepancy between the general assumptions and the counterintuitive findings on the Scandinavian excellence in aggregate health. To clarify the focus of this thesis, this thesis tentatively distinguishes between ‘the first Scandinavian health puzzle’ and ‘the second Scandinavian health puzzle’. The first puzzle indicates the generally accepted but counterintuitive Scandinavian underperformance in narrowing health inequalities, and the second puzzle is the questionably ignored but still counter-theoretical Scandinavia’s worse than expected aggregate health. The focus in this thesis is firmly on this questionably underexplored second puzzle.

In a similar vein, it should be also recalled that another theoretical assumption on the association between income inequality and aggregate health, often called 'Wilkinson Hypothesis', has long been challenged (Judge, 1995; Bobak et al., 2000; Wildman, Gravelle, and Sutton, 2003). In the systematic review on the association, Lynch et al. (2004), which is called "probably the most comprehensive independent... systematic review of the evidence” (Mills, 2012) conclude that there is "little support” for the relationship between income inequality and aggregate health. The conclusion, again, contradicts the theoretical assumption that narrow income inequality would enhance aggregate health. This contradiction is closely related to the second Scandinavian health puzzle.

In a bid to shed light on these confusing mix of assumptions and findings on the relationship between income inequality, welfare regimes and aggregate health, which are the two research questions of this thesis, this chapter gathers and reviews previous review articles.

This chapter consists of four parts. The first part briefly introduces the method of 'review of reviews' and its 'data': six previous (systematic) review articles. The second part conducts the ‘review of reviews’ for two sets of review articles respectively and analyses reasons the review articles on the identical subject produce contrasting conclusions. The third part discusses implications of the findings from the two rounds of review of reviews. The fourth and final part identifies ‘the second Scandinavian puzzle’ and also discusses the methodological and theoretical issue on systematic review for further studies. In the end, this 'review of reviews’ recognizes 1) the lack of a consensus on the Wilkinson Hypothesis, 2) the presence of the second Scandinavian puzzle and 3) methodological limitations of the previous (systematic) reviews.

3-2. Method

The method of critically reviewing previous systematic reviews on similar research questions has different names and, consequently, different definitions and connotations. Becker and Oxman (2008) use the term of ‘overview of reviews' defining it as "review defined to compile evidence from multiple systematic reviews of interventions into one accessible and usable documents” (p. 607). Gough et al. (2012) regard ‘review of reviews’ as “a systematic map and/or synthesis of previous reviews” (p. 49). Caution must be exercised as the methods have been mostly used in clinical medical research where systematic reviews on similar subjects have been constantly published with the accumulation of updated primary articles.

Petticrew and Roberts (2008)'s *Systematic Reviews in the Social Sciences* is a rare case of expanding the systematic review method out of the clinical research. The authors justifiably claim that “the science of systematic reviewing for social policy purposes is still relatively young” (p. xiv). It is not surprising that even they miss these methods of 'review of reviews' or 'overview of review' in their list of review methods. It means that this chapter aims to contribute to the development of this nascent area of the social policy literature by reviewing previous systematic reviews. Given that we have hundreds, or probably thousands, of studies on the relationship between income inequality, welfare regimes and health, it is worth overviewing the academic landscape with the wider perspective with this new methodological tool. In addition, there are relatively few systematic review articles for this subject. As can be seen in the following part, when this researcher strictly confines the search to 'systematic reviews,' only two journal articles can be collected, with which conducting review of reviews could be meaningless and impossible. That is why this researcher adopts Gough et al. (2012)’s concept of 'review of reviews' as it covers not only systematic review but also general review articles. The other term of ‘overview of reviews’ is much more commonly used in clinical medical research and has more detailed protocols (Higgins and Green, 2008; Smith, Devane, Begley, and Clarke, 2011). However, it may not be applicable to this social science research without radical modification.

As this researcher broadens the search to encompass review articles, four review articles can be collected in addition to the two systematic review articles. This chapter carries out a review on these six review articles. In addition, Gough et al. (2012) provides little more than a brief definition of ‘review of reviews’ and this chapter has no choice but to serve as a 'very rough review of reviews' rather than a 'systematic review of systematic reviews' at least until we have a more robust definition and refined guidelines to conduct ‘review of reviews’ adaptable to social science.

3-2-1. Article Selection

A primary article is to 'a systematic review' what a systematic review is to 'an overview of review' (Becker and Oxman, 2008). In other words, 'data' for a systematic review would be previous primary articles, and data for a review of reviews would be previous review articles.

To find relevant 'data', three steps were taken. At first, two online datasets (Pubmed and Web of Science) were searched. Falagas, Pitsouni, Malietzis, and Pappas (2008, p. 341) claim that Pubmed and Web of Science are two of the four most popular databases among researchers together with Scopus and Google Scholar. In particular, the use of Web of Science in systematic review is regarded as “the standard” (p. 341) in systematic reviewers.

The review was carried out in in the first week of November in 2015, with key words of ① 'welfare regime', ② ‘welfare state’, ③ 'welfare capitalism', ④ 'income inequality', ⑤ 'income distribution', ⑥ 'population health', ⑦ 'aggregate health', ⑧ 'health inequality', ⑨ ‘mortality’, ⑩ ‘life expectancy’, ⑪ ‘systematic review’, and ⑫ ‘review’. The articles need to have at least one of ①, ②, ③, ④ or ⑤ plus at least one of ⑥, ⑦, ⑧, ⑨ or ⑩ together with either ⑪ or ⑫. For example, one article can be eligible if it has this set of ① 'welfare regime', ⑤ 'population health' and ⑨ 'systematic review' but without any of the three, it would be filtered out. When this researcher confines articles with a key word of ‘systematic review’, only 36 articles are collected from the two databases. Another key word of ‘review’ could fetch 943 articles. Altogether, 979 articles are selected in the first step. Duplicate references are found by both databases.

The second step is to filter the searched articles with inclusion criteria that it should 1) be a review article published in an English-language peer-reviewed journals since this new century (2001), 2) systematically review primary articles whose subject is the relationships between either welfare regime or national income inequality and aggregate health in a cross-national perspective as it corresponds with this thesis' two research questions.

There are two reasons behind the choice of the publication period since 2001. At first, the dramatic change of human health indicators over decades (see Regidor et al., 2011) requires relatively contemporary data if the research foci are not time-variant trends. Second, for the last decades, refined and updated data have kept emerging, including Luxembourg Income Study or the Human Mortality Database, making 20- or even 10-year-old data seemingly obsolete. Lynch et al. (2004) point out that since only nine years ago (1995) from the timing of the writing, most of studies began to present different perspective from the previous ones due to “using better quality data” (p. 48). This study encompasses the longer 15 years (2001~2015) for the reviews. This process sifts out six relevant review articles including two systematic reviews and four reviews.

3-2-2. Two Rounds of Review of Reviews (RR)

Of the six reviews, three articles (Macinko, Shi, Starfield, and Wulu, 2003; Lynch et al., 2004; Wilkinson and Pickett, 2006) analyse the relationships between income inequality and aggregate health in industrialized nations. The other three (Muntaner et al., 2011; Brennenstuhl et al., 2011; Bergqvist, Yngwe, and Lundberg, 2013) examine relationships between welfare regimes and aggregate health in rich countries. It needs to be noted that the six review articles are the same in that they focus on the cross-national difference in aggregate health, but the differences between the two groups are that the first sees income inequality as the main health determinant and the second concentrates on welfare regimes. Coincidently, or probably mirroring research trends during the period of the publications, each set of three reviews were published relatively simultaneously within only a three-year time span at its maximum, which is good for this comparative analysis of the review articles.

This chapter conducts two rounds of reviews of reviews (RR) for the two sets of three review articles. The two rounds both consist of the identical three steps. They will 1) identify the conflicting conclusions for the similar research questions 2) examine each review article's interpretation of their selected primary articles' conclusions, and 3) analyse reasons behind the contrasting conclusions from the identical research questions.

3-3. Findings

3-3-1. First Round of RR: Income Inequality and Aggregate Health

The three review articles have all common research questions. Macinko et al. (2003, p. 407) clarify that it “reviews published literature on the relationship between income inequality and health outcomes” while Lynch et al. (2004, p. 5)’s first sentence was “this article reviews … studies examining the associations between income inequality and health.” Wilkinson and Pickett’s (2006, p. 1768) first sentence also starts with a question of “Whether or not the scale of a society’s income inequality is a determinant of population health.” However, despite their almost identical focus, their conclusions are contrasting.

i) Conflicting Conclusions

Macinko et al. (2003), after reviewing 17 cross-national comparative review articles, remain sceptical in supporting the theoretically expected relationships between income inequality and aggregate health. According to their review, out of the 17 articles, 11 support the theory but six articles do not. After combining review outcomes of other within-nation primary studies that are more supportive to the theory[1], the authors conclude that “the relationship between income inequality and health is unclear (p. 407).”

Lynch et al. (2004) are even more sceptical on the relationships. After reviewing 28 studies on cross-national comparative health, their conclusion is that 16 support the theory while eight refute it with the other four not belonging to either category. "The

[1] Out of 28 studies, 22 nations (78.6%) support the theory but other six studies remain sceptical.

evidence suggests that income inequality is not associated with population health differences - at least not as a general phenomenon - among wealthy nations” (Lynch, 2004, p. 81).

However, Wilkinson and Pickett (2006), after reviewing 28 cross-national comparative studies, reach a contrasting conclusion that “a large majority suggest that health is less good in societies where income difference are bigger” (p. 1768).

The key relationships are concluded as “unclear” (Macinko, 2003, p. 407), having “little support” (Lynch et al., 2004, p. 5) or what “large majority suggest” (Wilkinson and Pickett, 2006, p. 1768). Their stances are not just different but conflicting. Given that they review a similar list of primary articles with the identical research question and their publication dates are not very different, it is a perplexing contrast. To analyse the reasons, we need to more closely examine how each review has assessed and interpreted the evidence.

ii) Different Lists and Interpretations

Table 3-1 shows three lists of primary articles selected by the three reviews. The three columns show lists of primary articles included in each review. On each line, a total of 39 primary articles published between 1979 and 2005 are included in at least one of the three reviews. Three different colours fill each cell with dark grey meaning supportive of the theory, light grey being neutral or mixed, and white negative. These colours reflect the review authors’ interpretation on the evidence presented in each included study.

Table 3-1. Three Systematic Reviews' Lists of Selected Primary Articles

Lynch et al (2004)	Macinko et al (2003)	Wilkinson & Pickett (2006)
Rodgers(1979)	Rodgers(1979)	Rodgers(1979)
Flegg (1979)		Flegg (1979)
Flegg (1982)	Flegg (1982)	Flegg (1982)
Steckel (1983)		Steckel (1983)
Weatherby, Nam and Issac (1983)		Weatherby, Nam and Issac (1983)
	Muller (1985)	
Pampel & Pillai (1986)		Pampel & Pillai (1986)
		Lester (1987)
Legrand (1987)	LeGrand (1987)	Legrand (1987)
Pampel & Zimmer (1989)		Pampel & Zimmer (1989)
Wilkinson (1992)	Wilkinson (1992)	Wilkinson (1992)
Waldmann (1992)	Waldmann (1992)	Waldmann (1992)
Wennenmo (1993)	Wennenmo (1993)	Wennenmo (1993)
	Wilkinson (1994)	
Duleep (1995)	Duleep (1995)	Duleep (1995)
Judge (1995)	Judge (1995)	Judge (1995)
Davey Smith & Egger (1996)		Davey Smith & Egger (1996)
	van Doorslaer et al (1997)	
McIsaac & Wilkinson (1997)	McIsaac & Wilkinson (1997)	McIsaac & Wilkinson (1997)
Judge, Milligan & Benzeval (1998)	Judge, Milligan & Benzeval (1998)	Judge, Milligan & Benzeval (1998)
Hales et al (1999)	Hales et al (1999)	Hales et al (1999)
Lee & Bankston (1999)		
Bobak et al (2000)	Bobak et al (2000)	Bobak et al (2000)
Marmot & Bobak (2000)		Marmot & Bobak (2000)
Lobmayer & Wilkinson (2000)		Lobmayer & Wilkinson (2000)
	Weich, Lewis and Jenkins(2001)	
Lynch et al (2001)	Lynch et al (2001)	Lynch et al (2001)
Mellor & Milyo (2001)	Mellor & Milyo (2001)	Mellor & Milyo (2001)
Elison (2002)		Elison (2002)
Pampel (2002)		Pampel (2002)
Gravelle, Wildman & Sutton (2002)	Gravelle, Wildman & Sutton (2002)	Gravelle, Wildman & Sutton (2002)
Wildman, Gravelle & Sutton (2004)		Wildman, Gravelle & Sutton (2004)
		Beckfield (2004)
Macinko, Shi & Starfield (2004)		Macinko, Shi & Starfield (2004)
		Drain et al(2004)
		De Vogli et al (2005)
		Pickett et al(2005)
		Pickett, Mookherjee & Wilkinson (2005)
		Ross et al (2005)
16 supportive (57.1%)	13 supportive (68.4%)	19 supportive (55.9%)
4 mixed or neutral (14.3%)	1 mixed or neutral (5.2 %)	9 mixed or neutral (26.5 %)
8 negative (28.6%)	5 negative (26.3%)	6 negative (17.6%)

It can be seen that their interpretation of the same articles differ in some cases (e.g. Judge, Mulligan, and Benzeval, 1998; Lynch et al., 2001). Out of the 39, 15 are included by all the three reviewers, 12 by the two and 12 by only one.

It can be observed that three bottom lines (calculated by this author based on each review's interpretation) were not basically different from each other.[2] Around 60 percent of the primary studies support the theory in all the three review articles, with the rate at 57.1% for Lynch et al. (2004), at 68.4 % for Macinko et al. (2003) and at 55.9% for Wilkinson and Pickett (2006). Similarly, around 20 percent (17.6%~28.6%) refute the association according to all the three reviews as well. The differences in the rates seem more insignificant as Wilkinson and Pickett (2006) reach the most supportive conclusion with the lowest ratio of the theory-supportive primary articles (55.9%). On the other hand, Macinko et al. (2003) remain cautious even though they interpret that the majority (68.4%) of the primary articles support the theory.

iii) Reasons for the Conflicting Conclusions

Three factors may account for the differences in the three conclusions. First, their main areas of focus differ somewhat. For example, Macinko et al. (2003) seem to be concerned with lack of consistency or unanimity on the subject, emphasizing "inconsistent" (p. 407), "mixed" (p. 432) and "varied" (p. 430) research outcomes even after measuring the highest ratio of supportive studies (68.4%). On the other hand, Lynch et al. (2004) pay more attention to the most recent research outcomes. They stress that "Most of the studies with negative or mixed results were conducted after 1995, presumably using better-quality data" (p. 48) and that the relatively 'new' studies

[2] Macinko et al. (2003; 431) explain that they reviewed 17 primary international articles without specifying individual studies. However, its list of articles in their appendix shows 19 studies conduct international comparative research. This article's table 1 follows the latter counting.

have failed to replicate their previous positive findings. These reasons lead them to a negative conclusion with regards to the theory that low income inequality is associated with better overall health profiles.

Wilkinson and Pickett (2006), after calculating the relative numbers of supportive and unsupportive articles, report that the majority of the primary articles support the theory. However, their method of counting the primary articles is different from the other two reviews. They omit the cases of what this chapter calls 'neutral or mixed' after labelling them as 'partially supportive' and only include the supportive and unsupportive articles to calculate the supportive article's ratio. With the decreased denominator, the proportion rises up to 78 percent. "Of those classified as either wholly supportive or unsupportive, a large majority… suggest that health is less good in societies where income differences are bigger" (Wilkinson and Pickett, 2006, p. 1768). It needs to be added that Lynch et al. (2004) cite Macinko et al. (2003) while Wilkinson and Pickett (2006) cite Lynch et al. (2004) but don't cite Macinko et al. (2003).

As the second reason for the contrasting conclusions, their criteria for selection of articles are different and, in some cases, questionable. For example, Wilkinson and Pickett (2006) include eleven international studies on relationships between income inequality and homicide. Homicide is clearly one of the factors deciding a nation's level of aggregate health. However, this problematizes the exclusion of other factors, for example, suicide or death from traffic accident, both of which are of greater statistical importance (Lukaschek, Erazo, Baumert, and Ladwig, 2012).[3] Wilkinson and Pickett

[3] Seemingly mindful of this possible criticism, they also indicate the number of the homicide studies in brackets beside the total number of the reviewed studies. This article deducts the homicide articles in making Table 3-1 for comparability with the other two reviews that do not include 'homicide articles' under their reviews.

(2006) also include Lester (1987) who admits both homicide and suicide into the analysis but omit other studies that only analyse association between income inequality and suicide (e.g. Fernquist, 2003). It is well established that homicide, among all the mortality related statistics, is strongly associated with income inequality and all the eleven 'homicide' articles reviewed by Wilkinson and Pickett (2006) support the association. Given this, their selection criteria might be seen to have decisively shaped the outcome of their analysis.

The selection criteria of the other two reviews are also somewhat questionable. Macinko et al. (2003) include Muller (1985) who analyses association between income inequality and ‘political violence’. They also review Doorslaer, Wagstaff, and Bleichrodt (1997) who study relationship between income inequality and ‘health inequalities’ not aggregate health. Lynch et al. (2004) also include one homicide article (Lee and Bankston, 1999) in their review list. Lynch et al. (2004) note that they would “not include those studies examining income inequality and homicide” (p. 21). Pampel (2002), examining relationships between income inequality and ‘smoking’ is also included by both Lynch et al. (2004) and Wilkinson and Pickett (2006).

These inconsistent selections seem in part due to unclear screening criteria. Macinko et al. (2003) note that they searched articles with key words ‘income inequality’, ‘health’ and ‘inequality’ but without clarifying their filtering guidelines.

Table 3-2. Conflicting Interpretations of the Same Articles

Three articles in focus	Key sentences in each abstract	Macinko et al. (2003)'s interpretation	Lynch et al. (2006)'s interpretation	Wilkinson & Pickett (2006)'s interpretation
Judge, Milligan & Benzeval (1998)	"we find very little support for the view that income inequality is associated with variations in average levels of national health in rich industrial countries" (p. 567)	negative	negative	partially supportive or mixed
Lynch et al. (2001)	"The associations that do exist are largely limited to child health outcomes and cirrhosis" (p. 194).	yes and no	negative	partially supportive or mixed
Pampel (2002)	"supporting the diffusion rather than the societal inequality explanation (p.35)."	not reviewed	negative	partially supportive or mixed

Lynch et al. (2004) state that they select studies on associations between income inequality and health but again without clear criteria. Wilkinson and Pickett (2006) note that they compile "reports of research on the relation between income distribution and measures of population health" (p. 1769) without justifying inclusion of homicide studies.

Third, in some cases, the reviewers' interpretations of individual article's conclusion differ markedly. As seen in Table 3-1, the three reviewers agree in their interpretation of the 27 articles reviewed by more than two reviewers. On the three cases (Judge et al., 1998; Lynch et al., 2001; Pampel, 2002), their interpretations are divergent. Table 3-2 shows the three articles' key sentences in each’s abstract and three reviewers’ interpretations.

It is noteworthy that Lynch et al. (2006) categorize all three articles as 'negative' while Wilkinson and Pickett (2006) interpret all the three as ‘partially positive or mixed’. Macinko et al. (2003) remain relatively neutral. Their interpretative tendencies correspond with each reviewer’s final conclusions. It is not this thesis’s interest to decide which side gives better interpretations. However, it seems certain that their interpretive tendencies influence their final conclusions or, possibly, vice versa.

Another important issue is a reviewer’s selection of a term. Wilkinson and Pickett’s (2006) expression of ‘partially supportive’ for the articles producing mixed conclusions might be questionable. It might give readers a misleading impression that the majority of studies support or at least ‘partially’ support the theory. For example, Lynch et al. (2001), who they categorize as ‘partially supportive’, in fact conclude that income inequality was strongly related with greater infant mortality but the association was reversed among those aged 65 and older. The expression of 'partially supportive' might not be a precise word for these mixed results. To be precise, both ‘partially supportive and partially disapproving’ or ‘simply mixed’ might be better expressions. Likewise, all the three systematic reviews may oversimplify complex and often multiple findings. These can be significant limitations of the three review articles and will be discussed further in Chapter 4.

In addition, the timing of publication may also influence reviewer's conclusion. In the case of Lynch et al. (2004), out of their seven most recent researches, five refute the theory while for Wilkinson and Pickett (2006), five out of their seven most recent studies support the theory.

3-3-2. Second Round of RR: Welfare Regime and Aggregate Health

In this second round of RR, again, the three review articles' research questions are almost identical. They clarify that each "examines the role of… welfare state characteristics on population health" (Muntaner et al., 2011, p. 946), "assesses empirical studies that use a welfare regime typology in comparative health research" (Brennenstuhl et al., 2011, p. 399) or "review this research… on welfare state characteristics and health" (Bergqvist et al., 2013, p. 1234). Again, their conclusions are contrasting.

i) Conflicting Conclusions

Muntaner et al. (2011) support the Scandinavian regime's excellence in enhancing aggregate health. After reviewing 31 primary cross-national studies, they conclude that "Social democratic regimes tend to fare best with absolute health outcomes" (p. 946). After assessing 17 articles, Brennenstuhl et al. (2011) are less supportive. They conclude that "some evidence supporting the hypothesis that the populations of social democratic regimes are in better health" (p. 399). The third group of reviewers, Bergqvist et al. (2013) remain skeptical. Based on review of 25 previous studies, their final verdict was “Results are diverse and contradictory” (p. 1234). Like the first round of RR, this second round of RR again sees the contrasting conclusions of the three reviews with similar research questions.

ii) Different Lists and Interpretations

Table 3-3 presents three different lists of primary articles covered by the three reviews. A total of 55 primary articles published between 1994 and 2013 are reviewed by at least one of the three review articles. The meanings of white (negative), light grey (mixed) and dark grey (positive) cells are same with those of Table 3-1. Additionally, at the first column, we have dotted cells, for which this researcher could not find how Muntaner et al. (2011) categorize each primary article’s conclusion. Muntaner et al. (2011) provide the ‘total score’ summing up the number of positives, neutrals and negatives but do not clearly indicate what categories some reviewed articles belong to. The ‘brightness’ of each cell depends solely on each reviewer’s original judgement on each primary article's conclusions.

Unlike the three review articles in the first round, the three reviewer groups in this second round present different evaluations over the tally of positives, neutrals and negatives. According to Brennenstuhl et al. (2011), 12 articles, more than 70 percent of the reviewed articles, support the theory, but Bergqvist et al. (2013) conclude that only around 40 percent (10 articles out of 25 articles) support the theory. Both Muntaner et al. (2011) and Brennenstuhl et al. (2011) note that they find only one negative article, but Bergqvist et al. (2013) identify six negatives.

Table 3-3. Three Systematic Reviews' Lists of Selected Primary Articles

Muntaner et al (2011)	Brennenstuhl et al (2011)	Berqvist et al (2013)
Lahelma & Arber (1994)		
Veehoven & Ouweneel (1995)		
Eola et al (1995)		
Whitehead et al (2000)		
Veehoven (2000)		
	Navarro & Shi (2001)	
Fayissa (2001)		
Conley & Springer (2001)	Conley & Springer (2001)	
Ouweneel (2002)		
	Theorell & Vogel (2003)	
	Navarro et al (2003)	
	Coburn (2004)	
Raphael & Bryant (2004)		
		Kunst et al (2005)
Menon-Johansson (2005)		
Bambra (2005)		
Bambra (2006)	Bambra (2006)	Bambra (2006)
	Navarro et al (2006)	
Dahl et al (2006)		
Zambon et al (2006)		
Muntaner et al (2006)		
		navarro et al (2006)
Nordenmark et al (2006)	Nordenmark et al (2006)	
Chung & Muntaner(2007)	Chung & Muntaner (2007)	Chung & Muntaner (2007)
Rostila (2007)	Rostila (2007)	Rostila (2007)
		Olsen & Dahl (2007)
Eikemo et al (2008a)	Eikemo et al (2008a)	Eikemo et al (2008a)
Eikemo et al (2008b)		Eikemo et al (2008b)
		Eikemo et al (2008c)
Klomp & Haan (2008)		
		Espelt et al (2008)
Lundberg et al (2008)	Lundberg et al (2008)	
Bambra & Eikemo(2009)		
Sanders et al (2009)		
Sekine et al (2009)		Sekine et al (2009)
Avendano et al(2009)		
		Sanders et al (2009)
Bambra et al (2009)		Bambra et al (2009)
Burstrom et al (2010)		
Farfan-Portet et al (2010)		
Grosse et al (2010)		
Karim et al (2010)	Karim et al (2010)	Karim et al (2010)
	kangas (2010)	Kangas (2010)
	Ferrarini & Norstrom (2010)	
	Norstrom & Palme (2010)	
	Granados (2010)	Granados (2010)
		Huijts et al (2010)
		Dragano et al (2010)
		Deeming & Hayes (2012)
		Harding et al (2012)
		Raphael (2012)
		Chuang et al (2012)
		Richter et al (2012)
		Van der Wel et al (2012)
		Ploubidis et al (2012)
		Popham et al (2013)
19 supportive (61.3%) 1 negative (3.2 %) 11 mixed or neutral (35.5%	12 supportive (70.6%) 1 negative (5.9 %) 4 mixed or neutral (23.5%)	11 supportive (42.3%) 6 negative (23.1 %) 9 mixed or neutral (34.6%)

iii) Reasons for Conflicting Conclusions

Again, three factors help to explain these divergent conclusions with the same research question and similar list of primary articles under reviews.

First, their foci are different. Just like Wilkinson and Pickett (2006) in the previous round, Muntaner et al (2011) also look at where the majority 'votes' are and comment that "population health differences across welfare state regimes found a positive association between welfare generosity and better population health (19 studies, 61.3%)" (p 954).

Brennenstuhl et al (2011) seem to base their conclusionon on more detailed observation at the pattern of the studies. They stress that the supportive conclusions come largely from some articles that "examined mortality measures (e.g. infant mortality rate, life expectancy at birth, etc) and included specific policy instruments in analytical models (e.g. extent of public healthcare coverage, public health expenditure, dual family earner policies, benefit generosity, etc)" (p 399). Therefore, they reach this cautious conclusion that, not the majority, but 'some' evidences support the hypothesis on aggregate health. It is noticeable that their bottom-line ratio of supportive articles (68.4%) is higher than that of Muntaner group's (57.1%).

Bergqvist et al. (2013) 'split' their conclusion. As seen in the table, they divide each article's conclusion into several minor conclusions. For example, Raphael's (2013) conclusion is that the Social Democratic welfare regime has the lowest infant mortality rates among welfare regimes but no obvious pattern was found between life expectancy and welfare regime. Other writers may categorize this study as positive or simply mixed.

However, Bergqvist et al. (2013) indicate both the conclusions in their appendix. After this detailed examination, they reach this relatively elaborate conclusion that East Asian and Scandinavian countries tend to have better records of life expectancy and infant mortality respectively than other regimes and no consensus can be found regarding morbidity. In the end, their final verdict on the relationships is "diverse and contradictory" (p. 1234).

Overall, Muntaner et al. (2011) emphasize the 'total score' in the lead up to the conclusion, while the other two reviewers relatively pay attention to detailed trends in the previous studies.

Second, their selection criteria are also different and in some case questionable. First of all, one can easily notice that most of articles are reviewed only by a single review article. Out of the total of 55 articles, only five are reviewed by all the reviewers and only nine by two of the three. The remaining 41 articles are reviewed only once. One reason might the different searching and selection strategies. Bergqvist et al. (2013) limit their selection to those articles published only since 2005, eight years before the article's publication. They also include eleven recent articles published after the other two reviews. This different time coverage may in part account for little overlapping in articles selected.

However, even if we limit the period to the three review articles' overlapping years between 2005 and 2009, still more than half of articles (14 out of 24) are included only by a single review. It is in part due to some questionable selection of articles by the reviewers, especially by Muntaner et al. (2011).

Table 3-4. Conflicting Interpretations of the Same Articles

Two articles in focus	Key sentences in each abstract	Muntaner et al.'s (2011) interpretation	Brennenstuhl et al.'s (2011)	Bergqvist et al.'s (2013)
Eikemo et al. (2008b)	"People in countries with Scandinavian and Anglo-Saxon welfare regimes were observed to have better self-perceived general health in comparison to Southern and East European welfare regimes" (p. 2281)	Not clear	positive	Negative
Eikemo et al. (2008c)	No comments on the subject in the abstract, but table 2 shows Scandinavians fail to top the list (p. 572)	positive	Not reviewed	Negative

Even though Muntaner et al. (2011) declare their research question as examining "the role of ... welfare state characteristics, on population health" (p. 946) on the first sentence of its abstract, they confusingly include three studies on health inequalities (Dahl et al., 2006; Muntaner et al., 2006; Bambra and Eikemo, 2009) along with 28 other studies on aggregate health.

Several other primary articles, which are not related to the research questions, are also included. They are, for example, studies on Sub-Saharan infant mortality (Fayissa, 2001), on HIV prevalence in 149 nations (Menon-Johansson, 2005) and on new welfare state typology (Bambra, 2005). Other debatable articles are on mental health (Nordenmark, Strandh, and Layte, 2006) reviewed by both the Muntaner group and the Brennenstuhl group, on happiness (Deeming and Hayes, 2012) reviewed by the

Bergqvist group, or on oral health (Sanders et al., 2009) reviewed by the Bregqvist group again.

Another thorny issue in the selection is whether to count seemingly 'salami slicing' articles as a single or multiple ones. In Table 3-3, four articles (Eikemo, Bambra, Judge, and Ringdal, 2008; Eikemo, Huisman, Bambra, and Kunst, 2008; Eikemo, Bambra, Joyce, and Dahl, 2008; Bambra and Eikemo, 2009) have the same main authors, use a same dataset, and use mostly same methods on similar set of sample nations. It is another point to be discussed how to include and count them in a systematic review.

Third, the three reviewers' interpretations of individual article's conclusions are different with each other. Some interpretations seem debatable. Table 3-4 shows the two articles that the reviewers provide different verdicts. There could be more articles which reviewers fail to reach a consensus, but, as stated, Muntaner et al. (2011) do not clearly provide information on their interpretations of most of articles' conclusions except for only 10 articles as seen in Table 3-3. In Table 3-4, it is notable that Bergqvist et al. (2013) tend to be negative in comparison with the other reviewers. Again, it is not this chapter's interest or aim to judge which side's interpretation is better.

3-4. Discussion

These two rounds of RR examine the theoretical aspect of 'the second Scandinavian puzzle' or 'the Wilkinson Hypothesis' as well as critique the review and synthesis methods employed.

The first round of RR demonstrates the lack of a consensus on the Wilkinson Hypothesis or on the relationship between income inequality and aggregate health. In fact, except for the eponymous reviewer (Wilkinson and Pickett, 2006), the other two reviewers are rather sceptical or at least neutral on the hypothesis.

The second round of review also shows contrasting conclusions on the relationship between welfare regime and aggregate health. Except for Muntaner et al. (2011), the reviewers remain sceptical or cautious on the association. These outcomes are in contrast with general assumption of Scandinavian outperformance in enhancing aggregate health in comparison with other welfare regimes (Hurrelmann, Rathmann, and Richter, 2010; Bambra, 2011; Mackenbach, 2011). They have approved the Scandinavian excellence in aggregate health especially in emphasizing the Scandinavian underperformance in narrowing health inequalities within its population. However, this review of reviews corroborates the broad disagreements on the issue of Scandinavian success in boosting aggregate health. Given that the Scandinavian welfare regime has narrow income inequality and an egalitarian welfare system, the inconsistent conclusions can signify another theoretical limitation. Therefore this chapter proposes 'the second Scandinavian puzzle' in addition to the first Scandinavian one.

On the method of systematic review, there have been some beliefs that the method would help shed light on where reliable answers lie based on combination of the conclusions of majority studies. For example, some commentators stress that "only the systematic review process is capable of helping to clarify where the answer really lies" (Shadish, 2006, p. vii) and "Generally, the results of a single study are not worth disseminating. Syntheses of the results of studies are the appropriate product of research endeavour" (Black, 2001, p. 278).

It is even more puzzling to see from this article's two rounds of RR that (systematic) reviews on the same research questions with similar lists of primary articles can reach such contrasting conclusions. However, it should not lead us to a hasty verdict of 'failure of (systematic) review.' It is just like sentencing a death penalty on individual primary study for producing conflicting conclusions on a same research question with a same dataset. What matters in this line of logic is to analyses what causes such differences rather than to dump the method itself. And this chapter analyses three reasons for the contradictory reviews at least as long as the selected six reviews are concerned. The three reasons are reviewers' 1) different focuses, 2) different (and sometimes questionable) selection criteria and 3) different (and sometimes questionable) interpretations for some primary articles' conclusions.

This analysis of the three reasons has two implications. First, we can ascertain that behind the conflicting conclusions between the reviews, there are questionable practices or probably mistakes involved. If we sort out the mistakes or questionable practices, we can have less variability among review articles' conclusions. Second, and more importantly, even if we straighten out the questionable practices and reduces some mistakes, we may face contradictory conclusions from even very similar systematic reviews as long as they have different focuses, different selection criteria and different interpretations. Unlike the problem of the 'questionable' practices, these differences could be arguably and at least in part inevitable as long as each author clearly explains and justifies their own focus, criteria and interpretation in their review.

3-5. Conclusion

These two rounds of RRs focus on the Wilkinson Hypothesis and 'the second Scandinavian puzzle' on the relationship between income inequality, welfare regimes and aggregate health and find the inconsistent conclusions. The two hypotheses are closely related because the main reason behind the general expectation for the Scandinavia's good aggregate health is their relatively equal income distribution (Hurrelmann et al., 2010). In other words, neither the hypothesised Scandinavian good health nor the Wilkinson Hypothesis is given solid empirical backing based on the two rounds of RR.

The RR has two methodological contributions. First, it is arguably the first attempt to conduct an RR over the previous (systematic) reviews in the cross-national comparative aggregate health studies. RR is a relatively common method in clinical medical studies (Becker and Oxman, 2008) but this method is new and not even mentioned in Petticrew and Roberts (2008)'s *Systematic Reviews in the Social Sciences*, a rare case of introducing the general review method out of the clinical research. Second, the RR reveals the three reasons behind the conflicting conclusions of the previous (systematic) reviews. They are 1) different focuses, 2) different selection criteria and 3) different interpretations for some primary articles' conclusions.

The RR has two theoretical contributions. First, it identifies and suggests the presence of the second Scandinavian health puzzle after analysing the conflicting conclusions of the previous systematic reviews. The second Scandinavian puzzle has been questionably overlooked and ignored especially in comparison with the first Scandinavian puzzle, which has drawn wide attention (e.g. Huijts and Eikemo, 2009; Popham, Dibben, and

Bambra, 2013). The assumption on the Scandinavia's good aggregate health has been so prevalent that researchers take the Scandinavia's good aggregate health for granted (e.g. Bambra, 2011; Mackenbach, 2011). This second puzzle needs to be under close scrutiny and will also be further examined throughout the following chapters.

Second, based on the inconsistent conclusions from the previous systematic reviews, the RR affirms the disagreements on the Wilkinson Hypothesis that income inequality would lead to worse aggregate health. Again, it should be also noted that the second Scandinavian puzzle and the Wilkinson Hypothesis is mutually related because the Scandinavian states are expected to show the best aggregate health mainly due to their narrow disposable income inequality.

For future studies, several points need to be addressed. At first, practical protocols seems to be necessary in social science research. However, it should not necessarily be same as those already designed for clinical medical research such as the Cochrane Collaboration's guideline (Higgins and Green, 2008). The guideline for natural scientists can surely provide cross-disciplinary insights as did for this study but may not be wholly applicable to social science studies. Second, reviewers need to clearly justify their research focus, article selection criteria and their interpretation of their selected primary articles. Third, reviewers again need to implement their selection and interpretation more strictly to avoid questionable practices. Fourth, a new review method may be needed to incorporate multiple findings in each single article. The decomposition method in the next Chapter 4 can be one of the solutions. Otherwise, the thorny dilemma of oversimplification of multiple findings may haunt future systematic reviewers.

CHAPTER 4 SYSTEMATIC REVIEW

4-1. Introduction

From the ‘review of reviews’ in the previous Chapter 3, we find that some mutually different characteristics and questionable practices in the previous reviews may account for contrasting conclusions despite the same research questions. Based on those findings, this chapter conducts a systematic review. To overcome limitations commonly found in the previous reviews as discussed in the previous chapter, this systematic review takes the following three steps.

First, it avoids oversimplification of each article’s conclusions if their findings are complicated or mixed. Second, consequently, this systematic review is relatively less concerned about counting and summing up the number of studies of which conclusions support the hypothesis or not. Third, it selects relevant articles and interprets their conclusions in clearer and more justifiable ways. Methodologically, it develops its own 'decomposition review method', which involves decomposing and reorganising all the independent variables, dependent variables, methods and data of primary articles under review. The methodological details would be elaborated in the third part of the following methods section.

This chapter consists of five parts. After this brief introduction, the method part discusses both the article selection process under review and the new decomposition review method. The next part deliberates the findings from the systematic review before we move on to the fourth discussion and the final fifth part of conclusions.

4-2. Methods

This section discusses, first, the guidelines for selection of relevant primary studies and, second, the process through which studies are actually filtered out or chosen. The third part elaborates what the decomposition review method developed for this study is and how it works for analysis of the selected studies.

4-2-1. Article Selection Criteria

The articles are selected under the following seven criteria to fit the research questions on the relationship between income inequality, welfare regimes and aggregate health.

The articles should

1) be an empirical study published in an English-language peer-reviewed journal.

2) be based on comparative cross-national statistics including at least three nations with one of the five Scandinavian nations (Sweden, Finland, Norway, Denmark and Iceland) or one of Northeast Asian OECD member nations (Japan and South Korea) together with at least two other nations from mutually distinctive regimes. Accordingly, at least three nations from at least three different welfare regimes need to be compared.

3) compare only developed nations over a certain threshold of per capita GDP or members of OECD.

4) examine aggregate health, not health inequality.

5) NOT gauge self-rated health (SRH) as an aggregate health indicator. This has been regarded as a valid barometer for aggregate health within a certain society or nation (Idler and Benyamini, 1997) and used by quite a few researchers (e.g. Jen, Jones, and Johnston, 2009). However, it has been criticised for causing misleading outcomes especially in a cross-country comparative health study context (Sen, 2002; Barford, Dorling, and Pickett, 2010). The Nobel economics laureate Sen (2002) simply regards SRH as "having severe limitations and can be extremely misleading" (p. 860) especially for comparative health studies. The subjective indicator can cause “major problems” (Rostila, 2007, p. 235) or “serious concerns” (Tapia Granados, 2013, p. 139) for cultural and linguistic differences in the way people perceive and express their health condition. For example, elderly population over 65 in Japan, where the life expectancy is the highest in the world, report the worst self-assessed health among rich nations, while the subjective health shows one of the best outcomes in United States (Tapia Granados, 2013). This systematic review includes studies focusing on at least one non-survey-based aggregate health data such as mortality and life expectancy.

6) NOT examine only a partial indicator of aggregate health (a particular disease, obesity, oral health or smoking rate), analyse a cause-specific mortality like homicide rate, or cover data of only a certain unrepresentative group of the whole population, for example, the unemployed or ethnic minorities. However, studies on infant mortality rate,

child mortality rate or old-age mortality are eligible. Studies on either gender are also included.

7) be published since 2001. The two reasons for this guideline are same with those for the previous review of reviews in the previous Chapter 3.

4-2-2. Article Collection Process

i) Article Collection

The article searching process involves three following steps.

The first step was to search electronic databases of ‘Web of Science’ and ‘Pubmed’, which was carried out in the first week of November 2015. The key words are ① 'welfare regime', ② ‘welfare state’, ③ 'welfare capitalism', ④ 'income inequality', ⑤ 'income distribution', ⑥ 'population health', ⑦ 'aggregate health', ⑧ 'health inequality', ⑨ ‘mortality’ and ⑩ ‘life expectancy’. The selected articles need to have at least one of the independent variables (①, ②, ③, ④ or ⑤) and at least one of the five dependent variables (⑥, ⑦, ⑧, ⑨ or ⑩). Web of Science produced a total of 1,907 articles containing the combinations of the key words. They are all browsed with their titles and, if necessary, abstracts, and 297 articles are selected in the first round. Likewise, 214 articles are also selected from a total of 434 searched in the Pubmed database. Altogether, out of 2,341 articles searched in the two electronic databases, 511 articles were selected after a review of their titles and abstracts, with a number of duplicates.

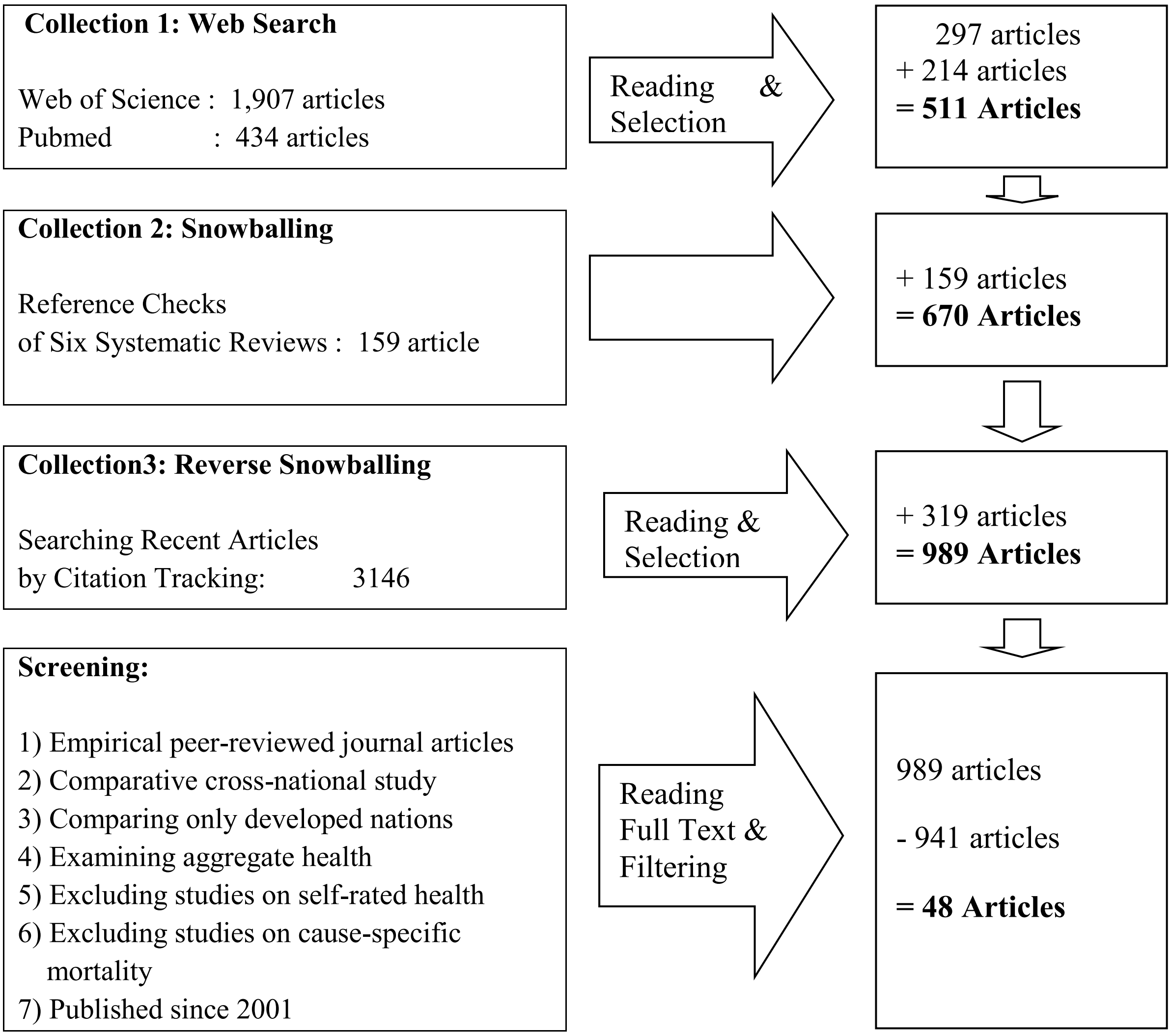
511 articles = 297 (Web of science) + 214 (Pubmed)

The second step is to browse lists of articles in the six systematic reviews that are discussed in previous Chapter 3. They each have lists of articles under their reviews and with all added up, a total of 159 articles are collected. Again there are a number of duplicates.

159 articles = 28 (Macinko et al., 2003) + 15 (Lynch et al., 2004) + 34 (Wilkinson and Pickett, 2006) + 31(Muntaner et al., 2011) + 17 (Brennenstuhl, Quesnel-Vallée, and McDonough 2011) + 34 (Bergqvist, Yngwe, and Lundberg, 2013).

The third step is to identify recent articles that cite the abovementioned systematic reviews or a few key primary articles. The key primary articles in the context are designated if they are reviewed by at least three systematic reviews of the six. In this way, eight key primary articles are selected. In turn, recent articles that cite the key eight articles are browsed in the Google Scholar search engine and selected after this reviewer read their titles, abstracts and main texts. For example, after reviewing the total of 234 articles that cite Mellor and Milyo (2001), this reviewer selected 24 articles. If the second step is 'snowballing' gathering skill, this third step is 'reverse snowballing' which is to find more contemporary articles by citation tracking (Sayers, 2007). This citation analysis of eight articles using Google Scholar fetches a total of 3146 articles. Of them, 319 articles are filtered through for the next round of review, as presented below.

Figure 4-1. Article Collection and Selection Process



- Six systematic reviews

Macinko, Shi, Starfield, and Wulu (2003) - cited by 133 articles, 9 selected.

Lynch et al. (2004) - cited by 640 articles, 41 selected.

Wilkinson and Pickett (2006) - cited by 999 articles, 38 selected.

Muntaner et al. (2011) - cited by 81 articles, 18 selected.

Brennenstuhl, Quesnel-Vallée, and McDonough (2011) - cited by 38 articles, 5 selected.

Bergqvist, Yngwe, and Lundberg (2013) - cited by 22 articles, 7 selected

- Eight key primary articles

Lynch et al. (2001) - cited by 384 articles, 52 selected.

Mellor and Milyo (2001) - cited by 234 articles, 24 selected.

Bambra (2006a) - cited by 85 articles, 20 selected.

Chung and Muntaner (2006) - cited by 136 articles, 20 selected.

Rostila (2007) - cited by 52 articles, 3 selected.

Eikemo et al. (2008b) - cited by 127 articles, 31 selected.

Eikemo et al. (2008c) - cited by 180 articles, 41 selected.

Karim et al. (2010) - cited by 35 articles, 10 selected.

The three-step reviewing process selects a total of 989 (= 511 + 159 + 319) articles. As mentioned, there are a number of duplicates. The following screening process involves reading of the articles' abstracts and full-texts and selecting some of them based on this study's seven criteria.

ii) Screening

Based on the abovementioned seven criteria, the following studies are excluded, 1) a book chapter (e.g. Theorell and Vogel, 2003), or a non-empirical article (e.g. Raphael and Bryant, 2006), 2) comparing regions within a single nation (e.g. Wilkinson and Pickett, 2008), only two nations (e.g. Olafsdottir, 2007) or only two welfare regimes (e.g. Tapia Granados, 2010), 3) covering developing nations (e.g. Fayissa, 2001; Drain, Smith, Hughes, Halperin, and Holmes, 2004), 4) examining only health inequalities, not aggregate health (e.g. Kunst et al., 2005; Borrell et al., 2009) 5) comparing self-rated health across the nations (e.g. Jen, Jones, and Johnston, 2009; Ploubidis, Dale, and Grundy, 2012), 6) examining not general aggregate health indicator but detailed health indicator such as level of smoking (e.g. Pampel and Rogers, 2004), obesity (e.g. Pickett, Kelly, Brunner, Lobstein, and Wilkinson, 2005), depressive symptoms (e.g. Dragano, Siegrist, and Wahrendorf, 2010) and happiness (e.g. Deeming and Hayes, 2012). 7) Studies published before 2001 were filtered out and excluded. Overall, hundreds of studies are omitted as a result of the screening.

Some articles that appear to violate one of the criteria 'survive' because they contain some findings relevant to this research question within their analysis. For example, Kondo et al. (2009) analyses cross-national differences in self-rated health but it also examines mortality. In this case, the latter finding remains for review. In a more subtle case of Harding et al. (2013) that compares mortality rates of employed and unemployed among the working age population in three different nations, this study only reviews its finding on health of the employed in the three nations. As seen, this review's fourth criteria exclude studies on health level of underrepresented group of populations. For the working age population, the employed are always the majority in all the developed nations (OECD, 2016a).

In other subtle cases, Navarro et al. (2006) and Safaei (2015) use relatively unconventional confidence interval of 90% in the lead up to their conclusions. However, to maintain consistency with other reviewed articles, this reviewer reinterprets their conclusions based on the more common 95% confidence interval. This long process of screening has left a total of 48 articles, of which list is presented in Appendix 4-4) together with their respective datasets, methods, cases, dependent variables, independent variables and conclusions.

4-2-3. Decomposition Analysis of Selected Articles

This chapter introduces a decomposition review method, which can be defined here as 'a systematic review method that decomposes the key components of each empirical study, namely independent and dependent variables, method and data to gather multiple findings from an individual article'. One of the key limitations of the six previous (systematic) reviews in the previous chapter is that all of them try to categorize primary articles in dichotomous or at best trichotomous ways. Even if a primary article conveys ample information in its findings on the relationship between income inequality, welfare regimes and aggregate health, it ends up categorised as either pro-hypothesis or anti-hypothesis and, at best, as an additional case of 'mixed'.

For example, Lynch et al. (2001) find that higher income inequality is strongly associated with greater mortality among infants, and more moderately associated with mortality among those aged 1-14 years in both sexes. For other generation groups aged 15~44, 45~64 and over 65, there was no statistically significant association between income inequality and mortality. Overall, life expectancy, arguably the most general health indicators, is not related to income inequality in a statistically significant way.

For these study outcomes, the three previous systematic reviews come up with different interpretations as seen in the previous section. Lynch et al. (2004) interpret the study as 'against the hypothesis', while the other two as 'supportive'. This researcher cannot decide which interpretation is better, especially when the same researcher (Lynch) reviews his own primary article. However, it is certain that they miss much of information in oversimplifying the delicate and dynamic relationships between income inequality and aggregate health. In another example, Beckfield (2004) finds that the relationship between income inequality and aggregate health is observed in the OLS regression model but disappears in the fixed effects model. This article is categorised just simply as 'anti-hypothesis' by Wilkinson and Pickett (2006).

These two examples of the 'monotone' review in fact carry significant implications for cross-national comparative health studies. The relationships between income inequality, welfare regimes and aggregate health vary depending on the methods and the health indicators used by researchers. It would logically follow that the choice of independent variables and data would influence their research outcomes as well. Therefore, even within a single study, the relationship could change dynamically. In this context, all the previous systematic reviews might have ignored these interactive dynamics between independent and dependent variables, methods and data, arguably because they are restricted by the 'one article=one finding' formula.

To incorporate the dynamics, this systematic review takes the three steps. The first step of this systematic review is to overcome the simple ‘one study=one finding’ formula and take into account the multiple findings in a single article. To take Lynch et al. (2004) as an example again, the review decomposes the study into five different findings as presented on Table 4-1, depending on its key components of health indicator (dependent variable) and datasets.

Table 4-1. Multiple Findings of Lynch et al. (2001)

<table>
<tr><th></th><th>Independent variables</th><th>Health indicators</th><th>Methods</th><th>Data</th><th>Support hypothesis?</th></tr>
<tr><td rowspan="5">Lynch et al. (2001)</td><td rowspan="5">income inequality</td><td>infant mortality</td><td rowspan="5">cross-section multivariate analysis</td><td rowspan="4">WHO Mortality Dataset</td><td>Yes</td></tr>
<tr><td>child mortality</td><td>Yes</td></tr>
<tr><td>working-age mortality</td><td>No</td></tr>
<tr><td>old-age mortality</td><td>No</td></tr>
<tr><td>life expectancy</td><td>WHOSIS Dataset</td><td>No</td></tr>
</table>

In the table, we can observe that a change in dependent variables, ceteris paribus, can result in contrasting outcomes except for life expectancy that uses a different dataset (WHOSIS). Likewise, this researcher decomposes the other 47 studies under this review depending on their number of findings they obtain. The decomposition is based on each article's choice of the four components[4]; ① independent variables, ② health indicators as dependent variables, ③ statistical methods and ④ datasets for health indicators. In the end, the decomposition produces 107 findings out of the 48 reviewed articles as seen on Appendix 4-4. If a research design is simple with a single independent variable and a single health indicator, its finding would be counted as one. Otherwise, there could be multiple for another article. Overall, each article has 2.3 findings on average ($2.3 \simeq 107/48$).

[4] There might be other components influencing the findings of a study. For example, datasets for independent variable, are much more various than other components. Due to space limit and efficiency of this review, other components are not analysed.

Table 4-2. Independent Variables, Health Measures, Methods and Data of Studies

	1	2	3	4	5	6	7	8
Independent variable	Welfare regime	Income inequality	Political tradition	Welfare expen-diture (genero-city)	Public health system	Wealth inequality	State	
Health measure	Infant mortality	Child mortality	Working-age mortality (14~64)	Old-age mortality	Life expec-tancy	All-age mortality	Life expec-tancy losses	
Method	descriptive statistics	bivariate analysis	Multi-variate analysis	Multi-level model	Decom-positon analysis	Time-series cross-section regression		
Health dataset	OECD health data	WHOSIS	WHO Mortality Dataset	Human Mortality Dataset	World Bank World Development Indicator	World Bank World Development Report	Individual nation's data	Other data-sets

The second step is to identify the pattern when using the four components in empirical studies under this review. For example, the reading of the 48 reviewed articles leads to a finding that all the studies used roughly seven health indicators: infant mortality rate, child mortality rate[5], working age mortality rate, old age mortality rate, life expectancy, all-age mortality and life expectancy losses. Likewise, the pattern of independent variables, statistical methods and datasets are reviewed and categorized by this researcher as presented in Table 4-2. Essentially, the table helps us to understand at a glance what types of independent variables, dependent variables, statistical methods and databases the previous empirical studies use.

[5] The under-5 mortality rates studied by some researchers (e.g. Collison, 2007; Chung & Muntaner, 2008) overlap with the first two indicators of infant mortality (0~1) and child mortality (2~13), and are categorised into the former in this study for the sake of analytical parsimony.

Table 4-3. Example of Coding: One Article Decomposed into Five Findings

	Independent variables	Health indicators	Methods	Data	Support hypothesis?
Lynch et al. (2001)	2	1	2	3	+1
	2	2	2	3	+1
	2	3	2	3	-1
	2	4	2	3	-1
	2	5	2	2	-1

The third and final step is to conduct 'coding' of each study. To take the example of Lynch et al. (2001) again, it is coded as '2' (income inequality) for its use of independent variable, as presented in Table 4-3. For its dependent variables, its five health indicators are coded as '1' (for infant mortality), '2' (child mortality), '3' (working age mortality), '4' (old-age mortality), and '5' (life expectancy), respectively. The study is again coded as '3' (multivariate analysis) for its method of correlation analysis with control variables.

For the last-column health dataset, Lynch et al. (2001) use the WHO's Statistical Information System (WHOSIS) for its life expectancy indicator (coded 2) and use the WHO Mortality database for all other indicators (coded 3). Moreover, if one finding is pro-hypothesis, it would be coded as '+1' and otherwise (anti-hypothesis) as '-1'.

Lynch et al. (2001) find that the income inequality and health indicators are statistically significantly associated only for infant mortality (coded as '+1') and child mortality (coded as '+1'), but not for the other three health indicators (all coded as '-1'). In this way, Lynch et al. (2001)'s analysis is coded into the five separate findings with mutually different combinations of independent variables, health indicators, methods, data and conclusions as shown in Table 4-3, which is a simpler form of Table 4-1.

Likewise, all of the 47 other articles are given either a single set or multiple sets of codes in accordance with their findings. The codes given each to the 48 articles' factors can be seen on Appendix 4-4. This rather complicated process of coding is necessary in the end because we have 107 findings to be analysed and the coding enables the researchers to process the data of the enlarged cases in a simpler way with a statistical program of R.

The coding process can lead to tabulation of 107 findings (from 48 articles) depending on their findings. For example, a two dimensional table with two axes of independent variables and health indicators can place all of the 107 findings in each relevant cells as seen Appendix 4-1. In the appendix, for example, we can see eight findings regarding the relationship between welfare regimes and infant mortality rate with five pro-hypothesis articles (Bambra, 2006a; Chung and Muntaner, 2007; Chuang et al., 2012; Raphael, 2013; Fritzell et al. 2015) and two anti-hypothesis articles (Karim et al., 2010; Regidor et al., 2011). Appendices 3-3 and 3-4 also outline the locations of all the findings on axes entitled 'methods versus health indicators' and 'data versus health indicators'.

This decomposition method, developed in this chapter, is original and unprecedented. This approach is different in three ways compared to conventional systematic reviews. First, this approach incorporates multiple findings in case a single empirical article contains more than one finding. Second, the incorporation process involves the decomposition of each reviewed article, if necessary, into multiple findings. The decomposition process depends on the use of the four components; independent variable, dependent variable, statistical method and datasets. In other words, the unit of analysis in this systematic review is not an individual article but an individual finding susceptible to change in any of the four components. Third, the enlarged number of cases (i.e. findings) necessitates coding and statistical processes rather than the conventional reading or vote counting of the reviewed articles.

This systematic review has the three contributions, closely related to its three characteristics. First, rather than compressing the otherwise ample information of each study into one monotone conclusion, this approach can make full use of the various, often, contrasting empirical findings even within an individual article. Second, the decomposition could provide a detailed look at how choice of independent variables, methods and data influences dependent variables. As seen Lynch et al. (2001), for example, given all the other components equal, a different dependent variable leads to a different finding. Third, the larger set of cases (i.e. findings) inevitably involves relatively complicated quantitative analysis but enables us to analyse the multidimensional interaction between independent variables, methods, datasets and dependent variables.

4-3. Findings

4-3-1. Two Dimensional Finding: Four Thresholds

Table 4-4. Number of Findings on Each Health Indicators [6]

	Infant mortality rate	Child mortality rate	Working-age mortality rate	Old-age mortality rate	Life expectancy	All-age mortality rate	Life expectancy loss	Total
Pro-hypothesis	32	6	2	2	9	2	0	53
Anti-hypothesis	13	1	6	5	24	4	1	54
Total	45	7	8	7	33	6	1	107

i) Age-threshold

In the previous part, this researcher splits the dependent variables of the studies into seven generational categories. All of the 107 findings from the 48 studies can be located along the seven dependent variables. If an article finds a significant association between an independent variable and a health indicator in hypothetically expected ways, it is categorised by this review as 'pro-hypothesis'. If not, it is categorised as 'anti-hypothesis'. The outcomes are presented in Table 4-4.

Given that the comparative health hypothesis expects that narrow income inequality or generous welfare system contribute to enhancing aggregate health in Scandinavia, the outcomes show an interesting pattern of ‘age-threshold’. In the table, the majority of findings support the hypothesis in infant mortality (32 articles out of 45 articles) and child mortality (6 out of 7), but the support radically diminishes from the working-age

[6] In the table, we can see 107 studies even though the total number of articles selected for the review is only 48. It is because of the way of counting research findings. If a study produces health indicators of, for example, infant mortality and life expectancy (often the conclusions are even contrasting), the study is counted as having two findings. In other words, in the table, the figures are the numbers of findings not those of the articles.

mortality (only 2 out of 8) and old-age mortality (2 out of 7). Even for the whole-life heath indicators, fewer articles support the hypothesis for life expectancy (9 out of 33) and all-age mortality (2 out of 5).

In total, the number of 'pro-hypothesis' (53) and 'anti-hypothesis' (54) findings is almost half by half, but the detailed observation reveals this age threshold over which the majority articles begin to show the contrasting patterns. Two articles (Lynch et al., 2001; Muntaner et al., 2002) analyse all of the generation-specific health indicators and the two reach the same conclusion, namely that: only the first two younger-generation health indicator outcomes support the hypothesis while the others not. Coburn (2004), the only reviewed article where findings are not coded in all the tables in this review as it does not indicate any p-values, an indicator of statistical significance, also finds the hypothetically expected relationship only under the age of 35. However, the relationship is reversed for the older generations.

The review outcomes suggest an age-threshold effect, which indicates the significantly different pattern of generational health influenced by socioeconomic factors. Based on the findings, it can be stated that the hypothetically expected impact of income inequality or welfare regime is strong for younger generation of infants or children under 14, but no longer statistically significant for older generations.

ii) Three More Thresholds: GDP Per Capita, Gender and Period

This systematic review could also identify some more thresholds over which some previous studies suggest the pattern of health indicators reverses or vanishes. The second and probably well-known threshold is the GDP per capita threshold. It is closely related to the Wilkinson Hypothesis that, in rich countries over a certain degree of income per capita, it is not economic growth but economic equality that drives up their population's health level (Wilkinson, 1992). For example, "economic growth is not a major factor" (Preston, 1975, p. 244) in increasing life expectancy or "at best, only a weak relationship between gross national product per capita and life expectancy (Wilkinson, 1992, p. 165). Over a certain threshold of the GDP per capita, its relationship with aggregate health would either weaken or disappear due to the theoretical curvilinear relationship between income and health. Among the reviewed articles, some studies support the GDP per capita threshold effect (e.g. Nowatzki, 2012; Pop et al., 2014) and some even contend that over the threshold, economic growth has even detrimental effects on infant mortality (e.g. Ferrarini and Norstrom, 2010) or longevity (e.g. Kangas, 2010). However, other studies (e.g. Babones,2008; Torre and Myrskylä, 2014) claim the persistent, significant relationship between income and aggregate health even over the threshold. The GDP per capita threshold issue remains open to further discussion.

The third threshold is the gender threshold. Out of a relatively small number of articles examine the relatively subtle difference between the gender, some observe the differences such as Torre and Myrskylä (2014) claiming that the female mortality rate for those aged over 65 is significantly related to income inequality but the old-age male mortality is not. Nowatzki (2012) also claims that female life expectancy is even more significantly related to wealth inequality than the cross-gender infant mortality rate, usually the most sensitive health indicator. Popham, Dibben, and Bambra (2013) find the East Asian women's relatively longer life expectancy than women elsewhere, while the Scandinavian males live longest among people from various types of welfare states.

On the other hand, in Lynch et al. (2001) and Muntaner et al. (2002), the gender threshold cannot be found. The term 'threshold' in this context may have different implications compared to other thresholds because the gender data is not gradual but nominal, but the possible presence of this between-gender difference in term of each gender group's vulnerability in health needs further analysis.

The other possible forth thresholds are 'the period threshold' over which time the hypothetically expected relationships between welfare regime, income inequality and aggregate health vanish for example from 'the early 21 century' (Regidor et al., 2011) or '1992' (Conley and Springer, 2001). Mello and Milyo (2001) also state that when income per capita is held constant, the Gini coefficient does not have a significant detrimental effect on aggregate health after the 1970s. The Social Democratic and Scandinavian countries, when compared with other welfare states, had the lowest infant mortality rates until the late 20th century, but the differences in infant mortality had been narrowed to be "negligible" (Regidor et al., 2011, p. 1187).

The proposed presence of the four thresholds calls for the more detailed analysis on the cross-national health study rather than the conventional monotone and oversimplifying approach. Among the threshold, due particularly to space limit, this review focuses on particularly the age threshold with the following three-dimensional approach. The gender and GDP thresholds are discussed and tested in Chapter 6, 7 and 8, but the period threshold, albeit identified in this chapter, is not tested in this thesis due to word limit. This limitation will be stated in Chapter 9.

4-3-2. Three-dimensional Findings

i) Relationships between Independent Variables and Health Outcomes

This systematic review could categorise the independent variables of the 48 articles into seven groups as seen on Table 4-5.[7] Several studies set more than one independent variable. For instance, Muntaner et al. (2002) use political tradition, income inequality and welfare state spending as the predictors. Consequently, the number of findings rises up to 107, more than double the number of the reviewed articles, as shown in the table. Apart from welfare regime and income inequality, the other five variables are also selected in this review, as they are all basically cross-national health comparative studies corresponding to this review's article selection guidelines.

Scandinavian nations which have a Social Democratic political tradition and thus almost synonymous with the Social Democratic welfare regime have more generous welfare states that contribute to narrow income inequality (Esping-Andersen, 1990, 1999; Korpi and Palme, 1998). The interrelated variables are expected to influence national-level aggregate health (Bambra, 2006a). Another variable of 'wealth inequality' is relatively new and unobserved in the international health comparative studies with only one exceptional study (Nowatzki, 2012). This review also includes this study for references, especially on the back of the recent growing interest in wealth inequality (e.g. Rowlingson and McKay, 2012; Piketty, 2014).

[7] This table shows only the number of findings in each cell. The appendix 4-1 is same as this table but contains all the names of articles in each cell.

Table 4-5. How Different Independent Variables Impact Health Indicators

	pro-hypothesis?	Infant mortality	Child mortality	Working age mortality	Old-age mortality	Life expec-tancy	All age mortality	Life expectancy loss	Total
welfare regime	pro	6							6
	anti	2	1	1	1	5			10
income inequality	pro	13	4	1		6	1		25
	anti	8		3	2	14	2	1	30
political tradition	pro	2	1						3
	anti			1	1	1	1		4
welfare spending	pro	5	1		2	2	1		11
	anti	1		1	1	1	1		5
public health system	pro	5							5
	anti	2				2			4
wealth inequality	pro	1				1			2
	anti								0
state by state	pro			1					1
	anti					1			1
total	pro	32	6	2	2	9	2	0	53
	anti	13	1	6	5	24	4	1	54

* The figure in each cell is the number of findings extracted from the total 107 findings

In the table, the majority (55) of findings examine the relationship between income inequality and aggregate health, and the verdicts on the relationships are divided roughly equally: 25 (pro-hypothesis) by 30 (anti-hypothesis). However, as found in the two-dimensional analysis, we can see the age-threshold again with the majority supporting the hypothesis for infant mortality or child mortality (17 against 8) while only one out of six is hypothesis-supportive for working age and old age mortality (1 against 5). For the general health indicators of life expectancy, all age mortality and life expectancy loss, fewer findings support the hypothesis (7 against 17). We can observe the age threshold effect because the relationship is supported by the majority of studies for infant and child mortality but not for health indicators of other generations.

For all the other independent variables, perhaps except for the last state-by-state comparison, the age threshold effect is observable. It can signify that the threshold is not due to any particular independent variable but this general trend can be seen all over the independent variables. The table also shows that the findings for the relationship between welfare regime and aggregate health are quite inconsistent. As seen in the last column and the second and third line of the table, out of 16 findings on the relationship, six support the Scandinavian welfare regime's better aggregate health records, but the other 10 do not.

With these discordant empirical findings, it is not surprising that the previous systematic reviews (Muntaner et al., 2011; Brennenstuhl et al., 2011; Bergqvist et al., 2013) on the research question reached contrasting conclusions. However, when this researcher divides the findings according to the different health indicators, the murky relationship becomes clearer. The Scandinavian welfare regime underperforms in terms of increasing life expectancy, as none of the five findings support the hypothesised Scandinavian good health as seen in Table 4-5. Six out of eight findings show that the regime does improve infant mortality rate.

All of the other independent variables discussed (income inequality, political tradition, welfare state spending, public health system) seem to have the hypothesised influence on reduced infant mortality rates with 26 out of the 37 findings supporting the relationship. When it comes to working-age or old-age mortality, the positive relationships are no longer supported by the majority of findings with only four out of 15 supporting the hypothesis. The situation is similar for the relationship with life expectancy and all-age mortality with 24 out of 33 studies refuting the hypothesis.

Table 4-6. How Different Methods Impact Health Indicators

	pro-hypothesis?	Infant mortality	Child mortality	Working age mortality	Old-age mortality	Life expec-tancy	All age mortality	Life expectancy loss	Total
descriptive statistics	pro	3		1					4
	anti					2			2
bivariate analysis	pro	4		1					5
	anti					2			2
multivariate analysis	pro	9	5			5			19
	anti	4		5	4	10	4		27
multilevel analysis	pro	2							2
	anti					1			1
decompositon analysis	pro								0
	anti				1	2			3
time-series cross-section	pro	14	1		2	4	2		23
	anti	9	1	1		7		1	19
total	pro	32	6	2	2	9	2	0	53
	anti	13	1	6	5	24	4	1	54

Ultimately, the Scandinavian welfare regime succeeds in reducing infant or child mortality rates significantly, but does not improve other health indicators. Across all of the other independent variables, arguably except for the last state-by-state comparison, the age threshold is observable. This could signify that the threshold is not due to any particular independent variable but that the effect is relevant for all of the independent variables.

ii) Relationships between Methods and Health Outcomes

As another three-dimensional approach, this review again categorises statistical methods of all the 107 findings into arguably six groups to analyse the relationship between statistical methods and health outcomes. The first statistical method is the simplest descriptive statistics between national or regime average health levels. The second method is bivariate analysis consists mainly of simple correlation analysis without control variables. The third method is multivariate analysis with some control variables such as GDP per capita (Lynch et al., 2001; Muntaner et al., 2002) or median wealth per capita (Nowatzki, 2012). The fourth multilevel analysis is used by two articles (Chung and Muntaner, 2007; Chuang et al., 2012). The method is designed to divide the variances in health outcomes into two levels of welfare regimes and individual nations.

The fifth decomposition methods is more concerned on the age-specific life expectancy loss (Shkolnikov, Andreev, Zhang, Oeppen, and Vaupel, 2011; Popham, Dibben and Bambra, 2013). Appropriate caution must be exercised, because the name of their method 'decomposition' is only coincidently same with the review method adopted in this chapter. The content of the two methods is completely different. While the first five methods focus on cross-sectional statistical methods, the last time-series cross-section (TSCS) analysis adds the chronological aspect to the spatial study (e.g. Chung and Muntaner, 2008; Engster and Stensota, 2011).

Table 4-6 shows that the methods do not seem to demonstrate significant differences in their impacts on health outcomes except for the decomposition analysis, which produces anti-hypothetical outcomes. In all the three decomposition analysis findings (one from Vaupel et al., 2011; two from Popham et al., 2013), the Scandinavian welfare states underperform in aggregate health outcomes but Japan and Southern European nations of Italy and Spain, traditionally welfare state laggards, record relatively good aggregate health outcomes.

However, when this researcher splits the health outcomes by the age threshold, different aspects of correlation or causation can be revealed. For infant and child mortality, the first three methods support the hypotheses (21 against 4). However, the last time-series cross-section analyses are less supportive of the hypotheses (15 against 10). On the other hand, for the general mortality rates of life expectancy, all-age mortality and life expectancy loss, the first three methods seem to be less supportive (5 against 18) than the time-series analysis (6 against 8). For the working-age and old-age mortality, no particular difference could be found between the methods. In other words, the age threshold effects are more often observed in the cross-section studies than in the TSCS ones.

iii) Relationships between Datasets and Health Outcomes

Table 4-7 illustrates that the use of different datasets may lead to different conclusions. After analysing the datasets used by the 48 articles (99 findings) [8], six popular international datasets could be identified as seen in the table. In addition, two articles (Ross et al., 2005; Harding et al., 2013) combine some nations' datasets from respective countries they examine. Other six articles (12 findings) combine more than two international datasets. Therefore, to analyse relationships between datasets and health outcomes, the first six datasets can be in particular under this review.

[8] This table has eight fewer findings that the previous table as Navarro et al. (2003) and Navarro et al. (2006) each with four findings do not specify what datasets they used.

Table 4-7. How Different Datasets Impact Health Outcomes

	pro-hypothesis?	Infant mortality	Child mortality	Working age mortality	Old-age mortality	Life expec-tancy	All age mortality	Life expectancy loss	Total
OECD health dataset	pro	11				2	1		14
	anti	1				3			4
WHOSIS	pro								0
	anti			1		4			5
WHO Mortality	pro	5	5						10
	anti		1	5	4		4		14
Human Mortality	pro	1	1			2			4
	anti	2			1	4		1	8
WB W. D. Indicator	pro	2				2			4
	anti								0
WB W.D. Report	pro					1			1
	anti								0
Combining national data	pro			2					2
	anti								0
Combining int'l data	pro	6			2				8
	anti	6				1			7
Other data	pro	3				2	1		6
	anti	4				8			12
total	pro	28	6	2	2	9	2	0	49
	anti	13	1	6	5	20	4	1	50

First of all, while most of datasets can be interpreted as containing the pro-hypotheis statistics for infant and child mortality, only the Human Mortality Database seem to have mixed data as two articles (Leigh and Jencks, 2007; Torre and Myrskylä, 2014) extract anti-hypothetical conclusions from the dataset.

Second, none of the international datasets were interpreted as having pro-hypothesis statistics yet for working-age and old-age mortality. Only some combinations of national-level or international-level datasets produce some pro-hypothesis health outcomes.

For the three general mortality indicators of life expectancy, all-age mortality and life expectancy loss, the two WHO datasets provide only anti-hypothetical health outcomes while the two World Bank datasets, on the contrary, provide the pro-hypothetical outcomes. Research based on OECD health dataset and the Human Mortality Dataset yield mixed conclusions. For these general mortality data, the six datasets show a clear contrast as far as the related findings are concerned. However, this relationship needs to be discussed with caution because the data is not the only factor dictating the health outcomes.

4-4. Discussion

The review of reviews (RR) in the previous chapter identifies the presence of the second Scandinavian puzzle on the regime's underperformance in aggregate health outcomes compared with the other welfare regimes. The systematic review in this chapter finds some clues to the puzzle.

First, the systematic review shows that the relationship between welfare regime and aggregate health are quite inconsistent, but the decomposition review method demonstrates that the Scandinavian welfare regime shows worse-than-expected outcomes in life expectancy in comparison with all the other welfare regimes in all the five studies as seen in Table 4-5. The welfare regime excels only in reducing the infant mortality rate as six out of the eight findings support its outperformance. These age-threshold effects have been invisible before, because the generation-specific patterns are not observed in the previous systematic reviews. While the Scandinavian welfare regime succeeds in reducing infant or child mortality rate significantly, little evidence is found for its relationships with any other health indicators.

Second, out of the other thresholds, the GDP per capita threshold may provide another clue to the second Scandinavian puzzle, because it can be a key factor in accounting for the contrasting conclusions among the primary articles. The GDP per capita threshold means “for rich countries to get rich add nothing to their life expectancy” (Wilkinson and Pickett, 2009, p. 6). Then there could be the ‘ceiling effect’ that for rich nations, a further increase in life expectancy might not be physically possible (Pop et al., 2013, p. 1040). Based on the claim, the converging trend of health indicators among rich nations over time (Lundberg et al., 2008) could at least in part explained away. In other words, the Scandinavian nations simply reached the 'ceiling' relatively early and could keep the gap only for a while that had gradually shortened and vanished. However, this account contradicts those arguing for ‘broken limits to life expectancy’ (Oeppen and Vaupel, 2002). They contend that “best-performance life expectancy has steadily increased by a quarter of a year per year, an extraordinary constancy of human achievement” (p. 1030) for the last 160 years. Therefore, the presence of the GDP per capita threshold and its impact on aggregate health can be another key point for resolving the second Scandinavian puzzle. In addition, the period and gender thresholds are observed, over which health outcomes show different patterns.

Third, as seen in the systematic review, the choice of methods may influence outcomes on the second Scandinavian puzzle. The analysis based on the simple descriptive statistics or the bivariate analysis, which examine the relationship between only two paired data sets, without any control variables may have limitations in untangling the complex array of health determinants. They, without doubt, provide insights to researchers for further analysis on the complicated web of pathways leading to aggregate health. However, without controlling for other variables such as GDP per capita or demographic characteristics, the simple approach may result in misleading conclusions (see Beckfield, 2004). This thesis utilises as many health determinants, set up in Chapter 2, as possible to overcome these limitations in Chapters 5 and 6. In addition, the choice of datasets needs also to be clarified because it may impact research outcomes.

Fourth, again on the methods, another concern especially in comparing aggregate health of different nations with different socioeconomic, environmental and probably nutritional backgrounds is the unmeasured heterogeneity. Without incorporating the unmeasured heterogeneity, the cross-national comparative statistical analysis can be confounded. In that respect, the pooled TSCS methods are often designed to account for the heterogeneity may control for all shared period factors and time-invariant country-specific factors, making them arguably more reliable tools (Pop et al., 2014; Torre and Myrskylä, 2014). This thesis uses the pooled TSCS dataset in Chapter 5.

Fifth, the case selection is critical for validity and reliability of researches (Geddes, 1990; Kim, 2015). The hypothetically expected relationships between income inequality and aggregate health appear and disappear repeatedly solely depending on the selection of nations (de Vogli, 2004; Babones, 2008; Pop et al., 2013). Pop et al. (2013) test the relationships with a random combinations of 10, 16, 17, 18, 19, 20 and 21 countries drawn from a sample of 23 rich nations with a same dataset and find the tests produce inconsistent conclusions. The selected cases of each studies are, not surprisingly, very different because their criteria are different. The minimum average incomes, qualified to be cases, vary, such as 5000 dollars (Gravelle et al., 2002), 6000 dollars (Beckfield, 2004) and 20,000 dollars (Wilkinson and Pickett, 2009). These inconsistent case selections may in part account for the contradictive conclusions. Chapter 5 introduces a simple but theory-backed method to clarify the case selection process.

Sixth, another challenge regarding the case selection is the 'too small N' problem. As the analysis is limited to developed nations, the sample size can be no more than 20 or 30 nations (Ebbinghaus, 2005). Some have already voiced concerns over applying statistical methods such as multiple regression on such a small pool of nations (e.g. Shalev, 2007). However, even out of the 48 articles, one study goes as far as to use multiple regression on 11 rich nations (Lindstrom and Lindstrom, 2006). Regarding studies with this small sample size, Macinko et al. (2004) note that "it appears that the debate about the role of social inequalities on health is far from settled" (p. 281). The pooled TSCS analysis is the attempt to overcome the small N problem by multiplying the number of nations by the number of years examined. In addition, Ross et al. (2005) is another pioneering attempt to use the bigger selection of cases, 528 regions in five nations (Australia, Great Britain, Canada, Sweden and the United States). Likewise, if we can have detailed regional datasets, it could give a breakthrough for this small N problem. In Chapter 5 and Chapter 7, the 292 regional units from the OECD regional dataset are introduced and arguably address the small-N problem.

Seventh, given this small N problem, it appears that East Asian nations over a certain threshold of wealth have been constantly ignored except for Japan. It is probably because the comparative studies focus only on the conventional set of welfare states such as 17 Western nations plus Japan (e.g. Esping-Andersen, 1990: Ferragina, Seeleib-Kaiser, and Tomlinson, 2013) despite the rise of new welfare states such as South Korea. This is surprising considering that in the relatively small number of studies including more than one East Asian nations, their records are impressive. Japan has shown one of the best health outcomes in some of cross-national studies (Lundberg et al., 2008, Vaupel et al., 2011). East Asian nations do not record worse health outcomes than other welfare states (Karim, 2010; Chuang et al., 2012) or do record the better outcomes than others (Popham et al., 2013). However, for example, South Korea appears only five times (Collison, 2007; Kondo et al., 2009; Karim, 2010; Chuang et al., 2012; Avendano, 2012) out of the 48 studies under the review. This thesis identifies Japan and South Korea as forming a distinct East Asian welfare regime.

Eighth, all the studies under this review have the common limitations as ecological research, failing to effectively refute the argument that any association between income inequality and aggregate health can be only an 'artefact' as it is only a logical consequence of the curvilinear relationship between individual income and health (Gravelle, 1998; Jen et al., 2009a). For them, "spurious or artefactual correlation at population level between population mortality and income dispersion will always occur if the effect of individual income on the individual risk of mortality is smaller at higher incomes than at lower incomes" (Gravelle, 1998, p. 317). To take a very simple example, if a millionaire gives 100 pounds to a poor man, the possibly enhanced health benefits for the poor person would be bigger than the negative health effect on the rich person, which leads to the enhanced aggregate health. Some researchers have used multilevel modelling to shed light on the individual-level interaction between rising income and entailing enhancement in health (Jen et al., 2009a; Ploubidis et al., 2012) but their data have limitations as they are all based on individual subjective self-rated health, which is widely viewed as having limitation especially in cross-national comparative context (Sen, 2002; Rostila, 2007; Tapia Granados, 2013). Given this data limitation, the second best choice might be use of more detailed regional database as is done in Chapter 5 and Chapter 7.

4-5. Conclusion

This chapter has presented a systematic review on 48 primary articles published since 2001. The new decomposition method, proposed for this chapter's systematic review, splits individual articles into multiple findings based on the four components of independent variables, methods, datasets, and health indicators. Then this researcher rearranges the relationships between the three components and the health outcomes. The approach demonstrates the following findings. First, the majority of articles support the hypothetical relationships between income inequality, welfare regimes and aggregate health for the indicators of infant and child mortality, but not for working-age, old-age mortality, all-age mortality and life expectancy. Here we can observe the age threshold. Second, over most of different independent variables such as welfare state spending or public health system, the age-threshold is observed by the majority of articles. Third, the age-threshold is more apparent in the relatively simple statistical methods, not controlling for confounding factors such as unobserved heterogeneity than other refined statistical methods such as multiple regression or time-series cross-section models. Fourth, the use of datasets might also influence the research outcomes in some cases as some datasets produce consistent outcomes for particular health indicators. Fifth, in addition to the age threshold, the three other threshold effects are identified in this systematic review. They are income (GDP per capita), gender and period threshold effects. The hypothetically expected relationships between health determinants and aggregate health are claimed to reverse or vanish over each of the thresholds. Given these, the cross-national health comparative studies are asked to approach the subject in more multidimensional and more statistically refined ways rather than conventional monotone and oversimplifying manners.

Overall, this systematic review has three methodological contributions, owing much of them to the four suggestions of the RR in the previous chapter. First, it includes and counts multiple findings from an individual journal article to avoid oversimplification of their possibly various findings. Second, this systematic reviews decompose each articles into multiple findings based on their use of the four components; independent variable, dependent variable, methods and datasets. Third, this systematic review conducts a quantitative analysis after coding the four components of each of the 107 findings, which enables researchers to analyse the ‘data’ in multidimensional ways.

This systematic review’s findings also suggest three theoretical contributions. First, we can again confirm the presence of the second Scandinavian puzzle, this time after compiling the findings from the 48 empirical studies. Second, the detailed look at the puzzle could help identify the four thresholds on age, gender, income and period, over which health indicators show different patterns. The age thresholds effect is especially noticeable, because Scandinavian health records are relatively better only in infant and child mortality rates, but neither in old-age nor for all-age indicators. Third, we again affirm the wide disagreement on the Wilkinson Hypothesis. Again, the majority of studies support the hypothesis for younger generation’s health, but not for that of older generations.

CHAPTER 5 METHODOLOGICAL ISSUES

5-1. Introduction

This chapter is devoted to the account and justification of selection of methods, variables and related datasets in response to this thesis' research questions on the relationships between income inequality, welfare regimes and aggregate health. In other words, this chapter elaborates on the methodological choices of the forthcoming finding Chapters 6 and 7. The previous chapter 4 demonstrates that the conclusions regarding the relationship between income equality, welfare regimes and aggregate health vary according to selection of variables, datasets and methods.

In particular, choice of methods warrants caution in this cross-national comparative health studies focusing on the Wilkinson Hypothesis studies (on income inequality - aggregate health relationships) and the Scandinavian puzzle studies (on welfare regime - aggregate health relationships). Commentators highlight some limitations in this 'national aggregate' approach especially when compiling data of a small number of rich nations (Gravelle, 1998; Ebbinghaus, 2005). There are six complications in carrying aggregate-level comparative health studies.

The first issue is the possible 'ecological fallacy.' Some critics in particular emphasise that this has troubled previous research in national-level comparisons (Gravelle, 1998; Jen, Jones, and Johnston, 2009). They claim that the cornerstone of the argument on negative relationships between income inequality and aggregate health are due to a 'spurious aggregate relation' (Jen et al., 2009b, p. 643) or a 'statistical artifact' (Gravelle, 1998, p. 382) incurred by the curvilinear relations between individual income and individual health.

The second is the possible regional variance problem. This can be another version of the ecological fallacy. Within a nation, there could be regional differences in health as well as income, meaning that the national average would end up averaging out the regional variance. For example, southern Italian region of Campania has a disposable income per capita of 11,485 dollar, approximately half of 20,951 dollar in that of north Italian Bozano-Bozen provice (OECD, 2015d).

The third issue is the 'small-N problem', arguably the biggest threat to any serious cross-national statistical approach. Shalev (2007) elaborates the possible misleading conclusions especially linked to conducting multiple regression analysis with a limited number of cases. "Since no study of the OECD area can have more than about 20 cases… Multiple regression in effect places imaginary countries in some of these empty cells when it seeks out the best linear fit that can be generated for the data at hand" (p. 268). The 'small N' problem limits the statistical efficiency as the small sample size widens the variance of the estimators in econometrics design with limited 'degree of freedom' (Dougerty, 2011; Wooldridge, 2008).

The fourth concerns outliers in the already few cases of the rich nations. The issue of including or excluding the exceptional cases has long been debated. For example, Saunders (2010) contends that, in international health comparative studies, the US health indicator is too extreme and the nation should be treated as an outlier. He claims that Wilkinson and Pickett (2009) include the US case to supports their hypothesis but without the US, their hypothesis simply collapses.

Heterogeneity between the nations is the fifth issue. Among the seemingly homogenous rich OECD members, the heterogeneity problem remains. "(T)he OECD member states range from tiny Iceland to the 1000 times larger USA, from 'rich' Switzerland to four times 'poorer' Turkey (in terms of GDP per head). Given these large differences in population and economic resources, it may be misleading to analyse each case as equally important" (Ebbinghaus, 2005, p. 136). These are also logically related to 'unobserved variables' problem that also distort the statistical outcomes (Carmines, McIver, and others, 1981). For example, there might be less discussed health determinants such as climate, dietary habits, natural disaster or other cultural, historical or psychiatric characteristics.

Sixth and finally, there remains the historical contingency problem. If all the Scandinavian welfare states had merged into one nation decades ago, the regression line for the smaller '30' OECD nations with only one Scandinavian Union, would be much different (Ebbinghaus, 2005, p. 138). In another recent example, if Scotland had voted "Yes" in its referendum in 2014 and gone independent, there might be an additional OECD member in the future and a new regression line in the data would be drawn.

Overall, the six complications could be summarised as 1) ecological fallacy, 2) regional variance, 3) small-N, 4) outlier, 5) heterogeneity, and 6) historical contingency problems. The quantitative approaches in this thesis are responses to these imposing methodological challenges.

In response to the challenges, first of all, this thesis proposes an original case selection process, developed in this chapter. This case selection issue has not been reviewed in the previous review chapters, because it is extremely challenging to compare billions of sets of various welfare state cases. For example, if ten nations are to be selected from the current 34 OECD membership, the number of mathematically possible combinations (of the nations) is expressed in a mathematical term as ${}_{34}C_{10}$, equalling 131,128,140 sets of nations. The impossibility of comparing studies based on their case selections does not necessarily mean that the case selection could be overlooked. In fact, comparative studies on welfare states are criticised for underestimating or ignoring this critical issue (Ebbinghaus, 2005; Kim, 2015).

Critics have warned that 'selection bias' can lead to erroneous conclusion (Geddes, 1990; Hug, 2003). For example, Geddes (1990) takes the two examples of faulty inferences incurred from poor case selection. The first mistake is when any characteristic that the selected cases have in common is misinterpreted as a cause affecting dependent variables. The second is when the selected cases cannot represent the whole. Then erroneous generalisation is made by assuming that a relationship between independent and dependent variables within the selected cases mirrors the relationships in the entire population.

This chapter is divided into three parts. The first discusses the methodological choices for the pooled TSCS regression (for Chapter 6) and the second for the multiple regression with the cross-regional data (for Chapter 7). Each part has five subcategories. With the introduction at the beginning, the second part justifies the selection of each method. The third and fourth parts turn to selection process of independent and dependent variables in addition to related datasets. The fifth discusses the case selection process. Then the brief conclusion will follow before we move on to the next two finding chapters.

5-2. Pooled TSCS Data Analysis

5-2-1. Introduction

In the previous Chapter 4, the four kinds of potential thresholds were identified, including those of age, income, gender and period in relation to the health outcomes. Over each threshold, the hypothetically expected relationships between income inequality (or welfare regimes) and aggregate health reverse their patterns or vanish. For example, the 'age threshold' divides the age groups into the infant or child generation where the health determinant hypothesis works and the other older generations where the hypothesis stops working (e.g. Lynch et al., 2001; Muntaner et al., 2002). In other words, in their models, income inequality is related to younger people's health in statistically significant ways, but that relationship may not apply to older groups.

In the case of the second hypothesis on the 'GDP threshold', some researchers contend that economic growth does not contribute to driving up aggregate health over a certain degree of GDP per capita (Preston, 1975; Wilkinson, 1992, Wilkinson and Pickett, 2009). Likewise, across different genders (Nowatzki, 2012; Popham et al., 2013) and periods (Conley and Springer, 2001; Regidor, 2011), the health outcomes are claimed to show contrasting patterns.

Given the dynamics of the threshold effects, which complicate the analysis of pathways linking health determinants and aggregate health outcomes, we cannot simply reach a monotone conclusion on the relationships between income inequality, welfare regimes and aggregate health. Otherwise, the findings would be oversimplifying the complex relations and committing generalisation errors as reviewed in the previous Chapters 3 and 4. That necessitates the deeper analysis of the four individual thresholds in examining this thesis' research questions.

For this end, this chapter takes the following four steps. First, on the age threshold effect, the three different dependent variables are used: infant mortality rate, old age mortality and life expectancy. For the second threshold of income, GDP per capita would be one of independent variables to examine whether and how it influences aggregate health indicators even among rich nations over a certain threshold of GDP per capita. To test the third threshold of gender, at least one dependent variable (life expectancy or old-age life expectancy) is subdivided into female and male statistics. Infant mortality would be an exception in that context, as it does not contain gender subcategory. On the fourth and last threshold of period, this thesis uses the time-series cross-sectional (TSCS) data to analyse the chronological pattern of health indicators in 1995~2012. Apart from the 'period threshold', the choices in response to the other 'three threshold effects' in this pooled TSCS regression are same in the following multiple regression.

5-2-2. Methods

This thesis uses pooled TSCS data analysis by combining cross-sectional units of 26 OECD member states and 18 years over 1995~2012. (The reasons for selecting 26 nations out of the total 34 OECD membership and 18 years will be explained below) The use of ‘pooled data’ in this thesis has the following practical reasons (Podesta, 2002; Plümper, Troeger, and Manow, 2005).

First, “the place-time marriage of data” (Hicks, 1994, p. 169) can multiply the number of observations (= Units × Times), increasing the number of ‘cases’ dramatically. The large number of ‘nation-year’ observations allows researchers to relax the imbalance between too many independent variables and too few cases and makes it possible to reach a more accurate statistical estimation.

Second, the pooled TSCS analysis can examine the variation of both of time and space simultaneously. In cross-sectional arrays in a single time point, for example, the widening income inequality trend over the last decades cannot be captured. On the other hand, the single-unit time-series data cannot find the influence of a nation’s relatively time-invariant variables such as environmental factor. The two-dimensional pooled data can help encompass both the temporal and cross-national variability of variables.

Third, the pooling can help us take into account exogenous effects common to all units by controlling for time effects. For example, the financial turmoil in the late 2000s could influence health outcomes in most of developed nations in either direct or indirect ways. This temporal effect could be controlled for by, for example, adding time dummies to the model.

Fourth, the pooled TSCS analysis could reduce the omitted variable bias. For example, in this thesis, an intuitively important health determinant of a nation’s dietary pattern could not be operationalised because of the data unavailability. This omitted variable can distort the statistical estimation (Clarke, 2005). However, the bias could be decreased by, for example, adding unit dummy variables to control for each individual nation's unique characteristics, which is, in other words, heterogeneity.

To avoid confusion in terminology on the pooled data such as 'panel', 'cross-section time-series', or 'longitudinal', this thesis follows the definition by Beck and Katz (2004): "“panel” studies almost invariably have single digit T's (with 3 being a common value) while the comparative politics TSCS data sets we work with commonly have T's of twenty or more" (p. 3). Beck (2004) again uses the term 'longitudinal data' as encompassing both TSCS and panel data. As the number of period units is close to 20 in this thesis, the term of TSCS is used in this thesis rather than other terms.

It should be also noted that the pooled TSCS data also poses technical challenges because it violates several assumptions for the ordinary least squares (OLS) regression, the most commonly used multivariate data analysis method in social science. The econometric adjustment of the model in response to the violations is discussed further in Chapter 6.

5-2-3. Independent Variables and Data

In chapter 2, after reviewing the theories, it was noted that a total of twelve determinants of aggregate health from possible eight theoretical pathways could be operationalised to become potential variables. They are ① income inequality indicator (e.g. Kondo et al., 2009) ② income (e.g. Pop, Ingen, and Oorschot, 2013) ③ education (e.g. Doorslaer and Koolman, 2004), ④ occupation (e.g. Mackenbach et al., 2008), ⑤ alcohol consumption (e.g. Hoffmeister, Schelp, Mensink, Dietz, and Böhning, 1999) ⑥ smoking (e.g. Lawrence, Mitrou, and Zubrick, 2009), ⑦ dietary characteristics (e.g. Armstrong and Doll, 1975), ⑧ water quality (e.g. Benova, Cumming, and Campbell, 2014), ⑨ air quality (e.g. O'Neill et al., 2003), ⑩ public health expenditure (e.g. Regidor et al., 2011), ⑪ 'redistributive effect from taxes and transfers (e.g. Joumard, Pisu, and Bloch, 2012) and finally ⑫ decommodification index (e.g. Bambra, 2006b). Ideally, all of them need to be included and examined to avoid possible 'omitted variable bias' (Clarke, 2005) in this quantitative analysis. The omitted variables can inflate or deflate the coefficients of some of the utilised variables (Dougherty, 2011) and then distort the statistical outcomes. The critical obstacle however is the absence of relevant data. We need to find and include the maximum number of variables out of the twelve possible variables, as long as any reliable datasets are available.

First of all, as ① income inequality indicator, the Standardised World Income Inequality Database (SWIID) (Solt, 2014) is used. There are three reasons to use the SWIID among various international datasets on income inequality such as OECD dataset, Luxembourg Income Database or another indicator by the University of Texas Inequality Project. First, SWIID provides most of the indicator for all the OECD member nations in 1960~2012. On the other hand, the Luxembourg datasets have significant missing values over time, for instance, with only one-time data available for the key nations such as Japan (2008 only) and South Korea (2006 only). OECD cross-national income inequality data also contain too many missing values for its member states.

Second, SWIID divides market and disposable income inequality coefficients and provides the data for both, which the University of Texas dataset does not. The different effects of the two income distribution indexes on aggregate health are one of the focal points in this thesis. Third, the SWIID is regarded as an improved version of the older collections of income inequality datasets by enhancing cross-national comparability (Corneo, 2011; Bjørnskov, Dreher, Fischer, Schnellenbach, and Gehring, 2013). However, the dataset is not free of limitations because the SWIID, based on its imputation model, inevitably provides "plausible data but not sufficiently credible data" (Jenkins, 2015, p. 668). Consequently, it should be noted that there must be some costs for using the SWII. For example, the statistical outcomes using the dataset from the thesis could be biased and misleading. However, this researcher has few choices but the SWIID due to lack of internationally comparable income inequality datasets with few missing datasets.

The Gini coefficient in SWIID is located between 0 and 1 and the lower the index, the more equal income distribution is. In other words, the figure of ‘0’ means every household has the same income while ‘1’ indicates one household takes literally ‘all’. The market income Gini is with few exceptions higher than the disposable income Gini (i.e. the after-tax-and-transfer income Gini) because the rich’s market income is cut by taxes and the poor’s income is increased due to transfers.

The mathematical differences between market income Gini and disposable income Gini is defined as “the redistributive impact of taxes and transfer” (Journard et al., 2012, p. 4). The market income Gini indicator in SWIID will be ① market income inequality variable in this thesis. Second, the redistributive impact of taxes and transfer (= ‘the market income Gini’ – ‘the disposable income Gini’) will also be calculated from the

SWIID datasets and utilised as the shorter term of ⑪ 'tax and transfer effects'. In addition, the combined disposable income Gini will also be analysed, because "research comparing pre- and post-tax income inequality would be worthwhile" (Rowlingson, 2011, p. 13) in comparative health studies.

It is noteworthy that most of the cross-national comparative health studies use the comprehensive disposable income Gini statistics (e.g. Gravelle, 1998; Mellor and Milyo, 2001; Wilkinson and Pickett, 2009). However, the division of ① market income inequality (distribution) and ⑪ 'tax and transfer effects' (redistribution) would help examine the dynamics of labour market and welfare states (see Esping-Andersen, 1990).

In addition to the income inequality indicators, ② GDP per capita is another variable. It is calculated in terms of purchasing power parity in 2010 constant prices for global comparison over time in the OECD datasets. For the education indicator, ③ the school enrolment ratio for tertiary education in the UNESCO datasets is selected. There are seemingly better data to gauge a nation's educational attainment, such as 'Enrolment rate among 20~29' or 'Expected number of years in education' in OECD datasets (OECD, 2013d), but they either don't cover the statistics in the 1990s or have missing data for key nations including Japan.

Occupation is another key variable in accounting for cross-national difference in aggregate health. With few comparable datasets to represent the occupational gradient in each nation, ④ the unemployment rate of the World Bank dataset may serve as one of variable as the "exclusion from the labour market and the absence of paid work" (Bambra, 2011, p. 746) can aggravate individual health. For the health hazardous behaviours, ⑤ 'annual consumption of pure alcohol in litre per capita aged over 15' is used. It is available from the OECD health dataset. A further measure is ⑥ tobacco consumption such as OECD's 'annual consumption of tobacco items in grams per capita aged over 15', but the data of some nations such as Italy and Portugal were not available.

Public spending on health is the last variable that can be used in this thesis. World Health Organisation (2015) has datasets on "total expenditure on health as a percentage of GDP (A)" and "general government expenditure on health as a percentage of total expenditure of health (B)". If this researcher combines the two datasets, ⑩ "the general government expenditure on health as a percentage of GDP" (c) will be calculated as this function works.

The government health expenditure as a percentage of GDP (%) = (A * B) /100.

The following health determinants cannot be operationalised in this thesis mainly because of the relevant data unavailability. There have been studies on the following independent variables ⑦ dietary characteristics (e.g. Armstrong and Doll, 1975), ⑧ water quality (e.g. Benova et al., 2014), ⑨ air quality (e.g. O'Neill et al., 2003), ⑫ decommodification index (e.g. Coburn, 2004; Bambra, 2006b), but, relevant pooled TSCS data for these variables are not available. For example, World Health Organisation also develops ambient (outdoor) air pollution database for 1600 cities and 91 nations. The index is represented by annual mean concentration of fine particular matter (PM10 and PM 2.5 each meaning particles smaller than 10 and 2.5 microns). However, the database only covers the period from 2008 to 2013.

Figure 5-1. Seven Variables in Pooled TSCS Model

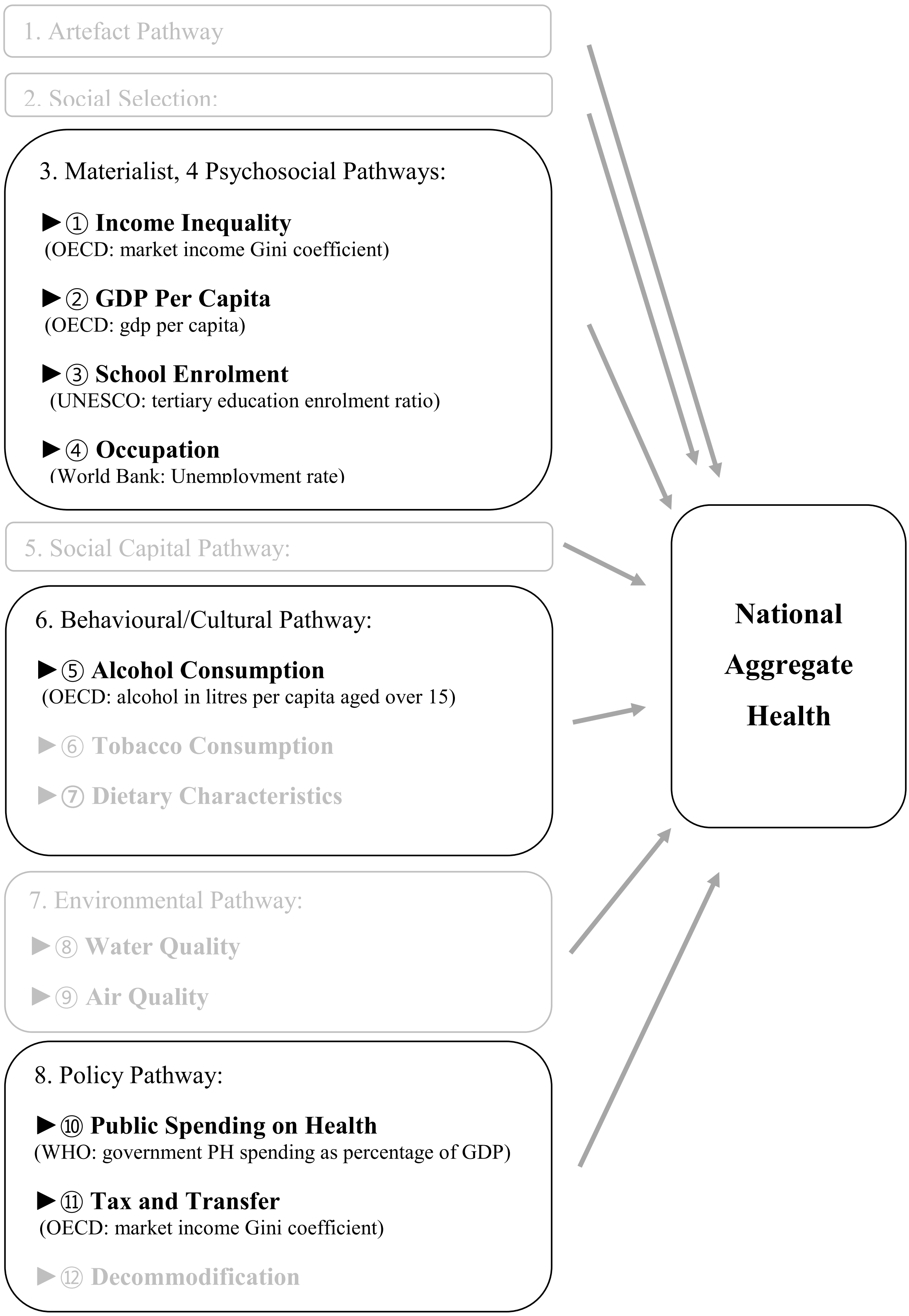


Table 5-1. Six Dependent Variables

	Definition	Available variables
Life expectancy at birth	"the average number of years that a person at that age can be expected to live, assuming that age-specific mortality levels remain constant" (OECD, 2014c).	① combined ② female ③ male
Life expectancy at 65		④ female ⑤ male
Infant Mortality Rate	"The number of deaths of children aged under one year of age that occurred in a given year, expressed per 1000 live births" (OECD, 2014c)	⑥ infant mortality

Given all the available variables, the methodological design in the cross-national study is presented as Figure 5-1.

5-2-4. Dependent Variables and Data

Of the four conceptual thresholds, the 'age and gender thresholds' require various dependent variables to test their presence. Six variables are presented in Table 5-1. The first 'life expectancy at birth' is to measure the lifetime health indicator while the second 'life expectancy at 65' is to gauge old-age mortality.

Infant mortality rate will serve as a younger-age mortality rate. Both the life expectancy at birth and at 65 contain respective female and male data to examine the gender differences in each indicator. The infant mortality rate data does not contain gender-related sub-datasets. This thesis avoids using morbidity data such as datasets on self-rated health (SRH), arguably one of the most common health indicators in cross-national comparative health studies (see Jen, Jones, and Johnston, 2009), because there are claims that the indicator has limitations especially in a cross-country comparative context (see Sen, 2002; Rostila, 2007). (On the limitations, see 4-2-1)

5-2-5. Case selection

This thesis aims to include all the 34 OECD member nations as long as the relevant datasets are available and related theoretical conditions are met. However, some nations are screened out due to following theoretical requirements. One of the key research questions in this thesis is directly related to the Wilkinson Hypothesis, positing that over a certain threshold of GDP per capita, it is not the average income but income inequality that determines aggregate health (Wilkinson, 1996; Wilkinson and Pickett, 2009).

However, critics point out the inconsistent and questionable case selection of the so-called "rich nations." For example, the Wilkinson group has been criticised for the omission of South Korea and the Czech Republic while including relatively poorer Portugal or the city state of Singapore (Saunders, 2010; Taxpayers Alliance, 2010). The case selection criteria have never been clear or justified (see Kim, 2015). Some set a certain level of GDP per capita as cut-off points such as around 5,000 dollars (Wilkinson, 1996; Gravelle, Wildman, and Sutton, 2002) or 12,500 dollar (Lindstrom and Lindstrom, 2006). Some use other guidelines for case selection such as the World Bank criterion (Ellison, 2002), OECD member nations but not World Bank-designated middle income nations (Macinko et al. 2004) or nations with available data (Lynch et al., 2001). Consequently, among the 48 articles reviewed in Chapter 4, the number of 'developed' nations ranges from 11 (Lindstrom and Lindstrom, 2006) to 40 (Vaupel, Zhang, and Raalte, 2011).

The different sets of developed nations in each study are found to be one of main reasons behind the different conclusions (De Vogli, Mistry, Gnesotto, and Cornia, 2005; Babones, 2008; Pop, Ingen, and Oorschot, 2013). For example, Pop et al. (2014), after conducting a simulation test for random combinations of 16~21 nations from a sample of 23 rich countries, find that the composition of the sample, ceteris paribus, decides statistical outcomes, leading to either statistically significant or insignificant associations.

Given this controversy, this thesis has taken a more systematic way to select the case nations. The starting point is none other than the Wilkinson Hypothesis, according to which the correlation between GDP per capita and aggregate health vanishes over a certain threshold of national income (Wilkinson, 1992; Wilkinson and Pickett, 2009). In other words, if this researcher tests the correlation between GDP per capita and life expectancy for the richest 10 nations, there is little correlation between the two variables. For example, people in Luxembourg, the richest nation in the world in terms of average income, live roughly one year shorter (80.7 years in 2010) than poorer Swedish people (the 10^{th} richest, 81.6 years). However, if this researcher tests correlation between the two variables for a wider set of 100 nations including multiple developing nations, the correlation is present: the richer, the healthier.

Consequently, there must be a threshold or a borderline over which the theoretical correlation begins to vanish or emerge if we use a medium number of nations. If we can find the threshold, then the nations over the threshold can be simply included for further analyses. On the other hand, countries under the threshold will be omitted. As a first step, a correlation analysis is conducted for all the 34 members of OECD, the correlation between the GDP per capita and life expectancy shows very strong statistical significance at 00.5 level ($t = 3.79$, p-value = 0.0006).

However, if the nations are omitted one by one from the poorest upwards, the t values gradually decrease (and p-values gradually increase) until the correlation is no more statistically significant. In the case of the 2010 data, the threshold GDP per capita is around 21,200 dollars dividing those of Estonia (21,056 dollars) and Hungary (21,477 dollars). It means that when the poorest five nations (Mexico, Turkey, Chile, Poland and Estonia) under the threshold are omitted, the correlation is no longer statistically significant. Then the rest 29 nations can be selected as the cases to be analysed. The data are from the OECD dataset. The method was Pearson's product-moment correlation analysis and the statistical software is R.

When this researcher conducts the correlation analysis repeatedly for the other years, we can see the list of the 'eligible' nations over each threshold to be included in each year as shown in Table 5-2. In respective years, the borderline between the white and grey groups of nations indicates the critical change in the statistical significance in terms of p-value at 0.05.

Over time, the thresholds can be observed increase gradually from around 13,300 dollars in 1995 (between of Slovenia and South Korea), around 14,000 dollars in 2000 (between Hungary and Czech Republic), around 17,000 dollars in 2005 (between Slovak and Hungary) and finally 21,200 dollars in 2010 (between Estonia and Hungary).

Table 5-2. OECD Nations in Order of GDP Per Capita

1995	2000	2005	2010
Luxembourg	Luxembourg	Luxembourg	Luxembourg
Switzerland	Norway	Norway	Norway
United States	United States	United States	Switzerland
Norway	Switzerland	Ireland	United States
Iceland	Netherlands	Switzerland	Netherlands
Austria	Iceland	Netherlands	Ireland
Denmark	Denmark	Iceland	Australia
Canada	Austria	Canada	Austria
Germany	Ireland	Australia	Denmark
Japan	Sweden	Austria	Sweden
Netherlands	Canada	United Kingdom	Canada
Belgium	Belgium	Sweden	Germany
Sweden	Australia	Denmark	Belgium
Australia	United Kingdom	Belgium	Iceland
Italy	Italy	Germany	Finland
United Kingdom	Germany	Finland	United Kingdom
France	Finland	Japan	France
Israel	France	France	Italy
Finland	Japan	Italy	Japan
Ireland	Israel	Spain	Spain
New Zealand	Spain	New Zealand	New Zealand
Spain	New Zealand	Greece	Korea
Greece	Greece	Israel	Greece
Czech Republic	Korea	Korea	Israel
Portugal	Slovenia	Slovenia	Slovenia
Korea	Portugal	Czech Republic	Czech Republic
Slovenia	Czech Republic	Portugal	Portugal
Hungary	Hungary	Hungary	Slovak Republic
Slovak Republic	Slovak Republic	Slovak Republic	Hungary
Poland	Poland	Estonia	Estonia
Estonia	Estonia	Poland	Poland
Chile	Mexico	Chile	Chile
Mexico	Chile	Mexico	Turkey
Turkey	Turkey	Turkey	Mexico

With the emerging economies especially in Eastern Europe, the ‘grey’ nations are growing in number. However, this researcher selects nations that have constantly been in the grey zone for the four timings to avoid ‘noises’ in further analyses. In other words, if a nation is not included in the grey zone at least once in the table, it is omitted. In the end, eight nations are excluded. They are five European nations (Slovenia, Hungary, Slovak, Poland and Estonia) plus Chile, Mexico and Turkey. The other 26 nations remain.

The next step is to categorise the 26 nations into the distinct welfare regimes. Since Esping-Andersen (1990) classifies 18 welfare states into ‘Three Worlds of Welfare Capitalism’ more than two decades ago, the controversy still continues on how many welfare regimes exist and which nations belong to which regimes. Despite Esping-Andersen’s (1999a) reluctance to add more to his original three welfare regimes, emerging welfare states in other regions require other types such as Southern European (Leibfried, 1992; Bonoli, 1997; Powell and Barrientos, 2004), East Asian (Kwon, 1997; Wilding, 2008), Central or Eastern European (Fenger, 2007; Aidukaite, 2009) and Latin American (Franzoni, 2008; Barrientos, 2009). This thesis encompasses all the discussed models in addition to the three original welfare regimes as long as the potential new regimes cover multiple welfare states.

Another unsettled and disputable issue is how to classify each welfare state to one of the multiple welfare regimes. To avoid subjective classification, this researcher can use one appendix table in Kim's (2015) review article where the 33 previous empirical studies are summarised and presented on how welfare states are classified into several regimes. Table 5-3 summarises the appendix, by which we can see which welfare regime each nation turns out to be classified into, according to the majority of the studies. The number in each cell in the table indicates the number of previous studies claiming the match between each nation and related welfare regimes.

Table 5-3. Regime Categorisation of 26 Nations in 33 Previous Studies

	Liberal	Conser-vative	Social Democratic	South European	Other Model	Total No. of Studies
United States	23				4	27
United Kingdom	20	1			10	31
Canada	17	4			5	27
Australia	18				9	27
New Zealand	13				10	25
Ireland	15	6	1	1	8	31
Switzerland	11	5	4		7	27
France		22.5	2	0.5	6	31
Austria		23	2		4	29
Belgium		21	5		5	31
Netherlands	2	14	5		10	31
Germany		22			9	31
Luxembourg	1	4				5
Finland		5	19		6	30
Denmark	1	2	25		3	29
Norway	1		24		4	29
Sweden			27		4	31
Iceland	1		1		1	3
Italy	2	15		6	7	30
Spain		3		6	6	15
Portugal	1	3		5	6	15
Greece		2		6	7	15
Japan		13	3		7	23
Korea					5	5
Israel					4	4
Czech Republic	1				2	3

For example, on the first line, United States was analysed in 27 previous studies, out of which 23 studies classify it into the Liberal regime and four other studies as ‘other models’ such as ‘productive welfare type’ (Hudson and Kühner, 2009).

By counting the number of studies on each nation, most nations can be judged where to be placed. Some thorny cases remain such as Italy, Japan, South Korea, Israel, Iceland and the Czech Republic. In the case of Italy, half of the studies, 15 out of all 30, regard it as belonging to ‘Conservative Regime’, but none of these 15 studies take into consideration the South European welfare regime. In fact, of six studies (e.g. Gallie and Paugam, 2000; Powell and Barrientos, 2004) that incorporate the Southern European model, all of them categorises Italy as the additional model. Consequently, Italy is categorised into the South European welfare regime in this thesis.

Japan is also one of the Conservative regime types according to the majority of the studies. Another East Asian OECD member state, South Korea, that has long been ignored in the comparative welfare state studies (Ebbinghaus, 2012; Powell and Kim, 2014), is categorised into ‘other types’ in the five studies that consider it. Given that the East Asian welfare model has drawn attention for its characteristics (Kwon, 1997; Aspalter, 2006; Wilding, 2008), this thesis categorises Japan and South Korea as forming a distinct East Asian welfare regime.

Israel and the Czech Republic are even more ignored cases as fewer studies include them. With some rare studies suggesting Israel as “Conservative” (Stier, Lewin-Epstein, and Braun, 2001) or “extended Mediterranean welfare states” (Gal, 2010), the nation seems to still wait for further analysis and in this thesis is tentatively categorised as not belong to any welfare regimes. The Czech Republic has been regarded as forming a distinct East European welfare regime together with other neighbouring ex-Communist East European nations such as Hungary and Slovenia (Fenger, 2007; Aidukaite, 2011). However, as the other East European OECD members are all screened out in the earlier case selection process, the Czech Republic remains the only East European welfare regime type. With no peers to be incorporated to the same category here yet, the nation also remains a single case not belonging to any of the other welfare regimes. Therefore Israel and the Czech will be analysed as an individual nation case, but not any member of the main welfare regimes.

The last case of Iceland has also received little attention with the only three previous studies all reaching conflicting conclusions. It is called “distinct” (Siaroff, 1994), “Liberal” (Saint-Arnaud and Bernard, 2003) or “Social Democratic” (Vrooman, 2012) as if to show its geographical location about half way between Scandinavia and America. This thesis however follows the conventional wisdom of Iceland as one of Scandinavian welfare model (Abrahamson, 1999; Kildal and Kuhnle, 2007).

5-3. Multiple Regression with Regional Dataset

5-3-1. Introduction

In the previous sections of this chapter, the six possible limitations of the cross-national ecological comparative health studies are discussed. They are 1) ecological fallacy, 2) regional variance, 3) small-N, 4) outliers, 5) heterogeneity and 6) historical contingency problems.

Table 5-4. Absence of Cross-regional Health Studies?

	Unit of analysis	
	Nation	Region (sub-national)
International perspective	▪ Wilkinson (1992) ▪ Conley & Springer (2001) ▪ Bambra (2006a) ▪Popham, Dibben, & Bambra, (2013) ▪ Torre & Myrskylä (2014), etc	?
Within-nation perspective	▪ Bartley, Sacker, & Clarke, (2004) ▪ Khang, Cho, Yang, & Lee, (2005) ▪ Chandra, Martinez, Mosher, Abma, & Jones (2005), etc ▪ Ford, Giles, & Dietz (2002)	▪Kennedy, Kawachi, Prothrow-Stith, & others (1996) ▪ Lynch et al. (1998) ▪Shibuya, Hashimoto, & Yano, (2002) ▪Blakely, Atkinson, & O'Dea (2003), etc

This researcher could arguably overcome some limitations in the previous pooled TSCS studies, but some issues still remain unresolved, such as the second regional variance problem. If the regional variance within a nation is overlooked, this researcher may commit another type of ecological fallacy especially when there is wide variance in health within a nation. The conventional welfare regime typology may commit such an ecological fallacy because the "debate has largely proceeded on the basis that coherent national welfare states exist" (Hudson, 2012, p. 455). Tuscany and Piedmont, the northern territory of Italy, is a noticeable example. In the nation generally categorised as either Conservative or South European welfare regime, the territory was long governed by Communist parties, which Navarro and Shi (2001) claim resulted in reduction in mortality rates through “a culture of solidarity and opportunity” (p. 486). In the national ecological approach, this characteristic case would be simply averaged out with other Italian regions.

Several studies have in fact focused on the regional differences but their focuses are confined to a single nation, in particular, such as United States (Kennedy, Kawachi, Prothrow-Stith, and others, 1996; Lynch et al., 1998) or Japan (Shibuya, Hashimoto, and Yano, 2002). There are few studies covering ‘regions over multiple nations’. Ross et al. (2005) are an exception, covering 528 metropolitan areas, but they are from only five nations including Australia, Canada, Great Britain, Sweden, and the United States. Such a ‘gap’ is even more striking when the research areas are categorised as seen in Table 5-4. The upper right cell is empty if we don’t have the exceptional case of Ross et al. (2005).

The absence of cross-regional comparative studies in international perspectives is mainly due to the lack of relevant comparable datasets. Ross et al. (2005) compile statistics from five individual national datasets. It is understandably challenging for researchers to assemble a comparable dataset from different sources. However, it is noteworthy that OECD has provided increasingly more detailed regional datasets (OECD, 2008; OECD, 2014d). The database contains key health indicators such as life expectancy, infant mortality rate and crude death rate in hundreds of regions in the OECD member states, enough to be utilised for quantitative analysis. The datasets have largely been ignored in the cross-national comparative health studies with no journal articles available online that proclaim use of the datasets. This thesis, for the first time for the purpose of international comparative health research, uses the OECD regional datasets of which details will be elaborated on in the next sections.

By using these detailed regional datasets, this thesis also overcomes the abovementioned six limitations of conventional cross-national ecological health studies. Of the six limitations, the first potential 'ecological fallacy' is perhaps the only one which the regional approach cannot provide a clear solution to, because it is also another form of regional 'average' (like the national 'average'). However, given the data availability, this is based on the most detailed 'objective' datasets, unlike those based on subjective 'self-rated health' ratings.

Regarding the second 'regional variance problem', it can be at least partially resolved by dealing with each sub-national region, not the whole nation, as an individual case. In the next findings chapter, significant regional variances within nations will emerge and be analysed. For example, some egalitarian Scandinavian regions turn out to have wider income inequality than those of some 'Conservative' regions.

The third small-N problem is resolved by enlarging the number of cases from around 30 nation states to 292 sub-national regions. Each region can be treated as a distinct case as each has its own data on education, income, income inequality, poverty and most importantly, health. It opens doors for more accurate statistical analysis.

On the fourth 'outlier' issue, if there are outliers, this researcher can simply exclude the extreme regions. In fact, in the small-N cross-national analysis, the USA is in most cases an outlier (Saunders, 2010; Taxpayers Alliance, 2010) with some extreme health outcomes. The nation poses dilemma, as it is too big to be ignored. However, as we have a relatively large number of cases, a couple of omissions would not hugely influence or change the results.

The most difficult challenge used to be the heterogeneity issue. In the case of different population sizes of the nations, the complication is addressed in this thesis, as OECD (2015e, p. 1) “defines regions as the first administrative tier of sub-national government” such as states in the United States, provinces in Canada, or régions in France. Additionally, homogeneity of GDP per capita also becomes higher as regions below a certain threshold were dropped in the following case selection process. However, it should be also pointed out that there might be “unobserved heterogeneity” (Wooldridge, 2008, p. 444) remaining unresolved.

The last ‘historical contingency problem’: the problem of 'what if Scandinavian countries are unified?' can be also resolved as the regional approach focuses on regions, not on nation states. In other words, even if the Scandinavian nations are unified, the regional cases for this analysis would remain the same as before the imaginary unification.

5-3-2. Methods

This researcher uses two statistical methods of ‘analysis of variance (ANOVA)’ and ‘multiple regression’. The OECD regional datasets have two characteristics. The first is that the regional datasets contain only one- or a few-time statistics for key variables. For example, the data for market income inequality and disposable income inequality are available only in 2010 not for any other years. Other independent variables such as air quality and education have also their data available only in 2000 and 2013. When combined, the statistics could be operationalised as just one-time model. This certainly has some limitations in generalising the finding over time.

The second characteristic, however, is that this cross-regional datasets contain statistics of hundreds of sub-national observations, enabling us to conduct the multiple regression, which was virtually impossible with the limited number of the national units and caused debates over its validity with the longstanding ‘small-N’ problem (see Esping-Andersen, 2007; Shalev, 2007; Scruggs, 2007).

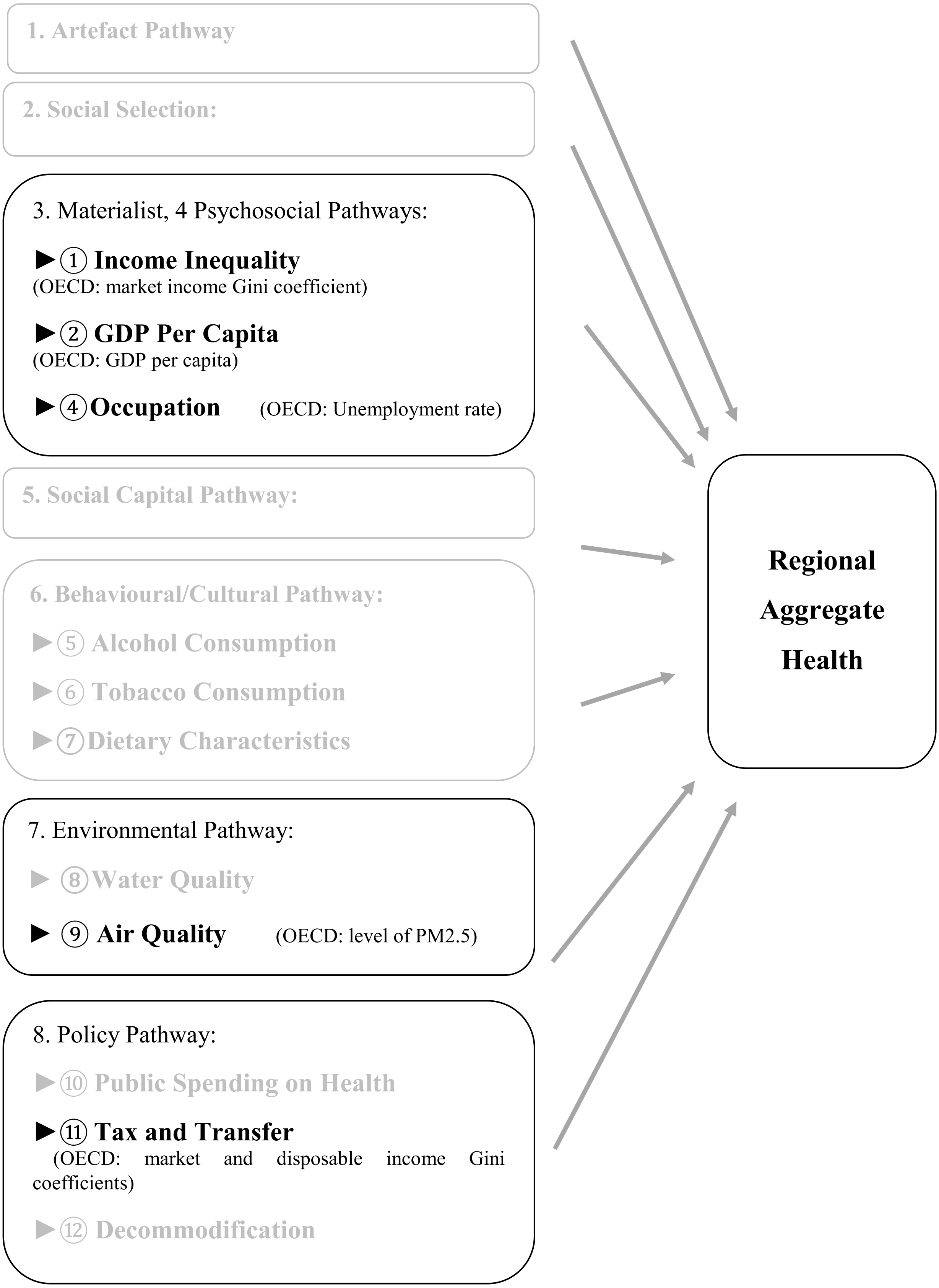
The two methods of multiple regression and ANOVA are conducted with five independent variables (income inequality, GDP per capita, occupation, air quality and ‘tax and transfer effects’) and six dependent variables (three life expectancy indicators for total, female and male population, infant mortality rate, crude youth death rate, old-age mortality rate). The outcomes of the analyses will be presented in Chapter 7.

5-3-3. Independent Variables and Datasets

The datasets are all from OECD regional database on 366 Territorial Level 2 (TL2) regions of 34 OECD member states. The TL2 region is “the first tier of sub-national government” (OECD, 2015e, p. 243) such as states in the United States, provinces in Canada, or régions in France.

In the regional database, This researcher finds six statistics that can be operationalised among ①~⑫ theoretically potential independent variables, identified in Chapter 2.

Figure 5-2. Five Independent Variables in Inter-regional Model



The first available statistic is ① the regional GDP per capita in 2000~2013 (in terms of constant prices, constant PPP) to gauge the regional income level. The second possible variable is ② income inequality indicator of market income Gini coefficient[9] in 2010. The third is ③ the 'share of labour force with at least secondary education' in 2000 or 2013, which is the only regional statistics related to education. The other available statistics is the air quality represented by ⑧ the level of 'pm 2.5' (the fine particulate matter in micrograms per cubic metres) in either 2000 or 2013. As the last variable, this researcher can again calculate the re-distributional ⑩ 'tax and transfer effects' as the database has the disposable income Gini in addition to the ② market income Gini coefficient. The 'tax and transfer effects' are calculated as the same ways as that in the cross-national design.

The other variables simply don't have matching statistics in the OECD regional database. The independent variables affecting the regional aggregate health can be presented as Figure 5-2.

5-3-4. Dependent Variables and Data

Just as the previous cross-national analysis is designed to use multiple dependent variables to test the 'age and gender thresholds', this cross-regional analysis also need as many variables as possible for the same reason.

- [9] Hungary's Gini coefficients are only available for the three larger geographical units of (Central Hungary, Transdanubia, Great Plain and North). Therefore, seven sub-regions are given one of the three Gini coefficient figures based on their locations. It is the same for the Polish smaller 16 using one of Ginis for the six bigger regions and for the Turkish 26 smaller regions for 12 bigger regions.

After reviewing the 12 available data in the OECD regional database, this researcher selects six statistics to use as dependent variables as presented in Table 5-5. The first ① 'life expectancy' is to measure the life-time health indicator with the gender group (② female, ③ male) data also available to test the gender threshold. To analyse the age threshold, three different age group data are also used including ④ infant mortality rate (aged 0~1) ⑤ crude youth death rate (aged 0 ~14) and ⑥ old age mortality rate (aged over 65). The gender statistics for the three data are not utilised in this thesis, because, in the case of the infant mortality rate, there are not sufficient regional observations (only 109 regions and more than 200 missing data). The 'gendered' crude youth death rate data are not also used here because it contains seemingly erroneous coding in the datasets. For example, the crude youth death rate (0-14 deaths for the 100,000 same-age group) in New South Wales in 2010 is 32.9 for females and 0.4 for males. Not only the single case, the OECD average crude youth rate in 2010 is 49.5 for females and 0.6 for males. It is highly likely that the female statistics are coded to be multiplied 100 times more than the male data, given that the combined average of the two genders in 2010 is 0.42. Then it is likely that the female data is 0.495 not 49.5, but this researcher cannot include the data without certainty.

The last old age mortality rate data is not present in the OECD database. This data is calculated for this thesis by combining two regional data of 'annual deaths of old age group over 65' and 'population of old age group over 65' in the OECD regional database. Then the old age mortality rate is calculated, following the definition of age-specific mortality rates "calculated by dividing the number of deaths registered in a particular age-sex group in a calendar year by the mid-year population estimates for that age-sex group" (Office for National Statistics, 2014).

Table 5-5. Six Dependent Variables

	Definition	Available variables
Life expectancy at birth	"the average number of years that a person at that age can be expected to live, assuming that age-specific mortality levels remain constant" (OECD, 2014c).	① total ② female ③ male life expectancy at birth
Infant Mortality Rate	"The number of deaths of children aged under one year of age that occurred in a given year, expressed per 1000 live births" (OECD, 2014c)	④ infant mortality rate
Crude youth death rate	"0-14 deaths for 100 000 population of same age group" (OECD, 2014c)	⑤ crude youth death rate
Old age mortality rate	OECD regional dataset does not contain this measure. It is calculated, following the age-specific mortality calculation (Office for National Statistics, 2014)	⑥ old age mortality rate

The old age mortality rate (%)

= (Annual death of old age group)/(Population of old age group)*100

The calculated outcome may serve as 'the old age health indicator' but is apparently NOT the official OECD data. Therefore, this researcher would not go further to calculate their gender subgroup data.

5-3-5. Case Selection

Out of the total number of 366 TL2-level regions in the 34 OECD member nations, at first the five Estonian regions should be omitted as they lack most of key statistics such as income inequality and all of health indicators. In the next step, with the remaining 361 regional cases, this researcher can conduct a correlation analysis between regional GDP per capita and their life expectancies just like this researcher has done for the national-level case selection in the earlier part of this chapter. One of the key research questions in this thesis is related to the Wilkinson Hypothesis, positing that over a certain threshold of GDP per capita, it is not the average income but income inequality that determines the aggregate health (Wilkinson, 1996; Wilkinson and Pickett, 2009).

By omitting regions upwards from the bottom of the regional income table, the threshold can be reached, over which the correlation between GDP per capita and life expectancy is no more statistically significant. This method can justify the omission of some relatively poor OECD regions based on their GDP per capita lower than the threshold. It needs to be recalled that the focus of this thesis is relatively wealthy regions or welfare states over the threshold. If the correlation analysis is conducted with all the available 361 cases, the correlation turns out to be very strong ($t = 9.7$, p-value ≈ 0), implicating the strong impact of income on health. If we tentatively place the threshold of GDP per capita at 10,000 dollars, excluding 18 relatively poor regions and conduct the correlation analysis again, the correlation is still fairly strong ($t=8.0$, p-value ≈ 0). After raising the threshold up to 17,300 dollars, the correlation is finally no more statistically significant at 0.05 level ($t=1.92$, p-value $= 0.055$). The method was Pearson's product-moment correlation analysis. The statistical software is R.

Table 5-6. 292 Regions, Nations and Welfare Regimes.

Regimes	Nations	regions	Regions
Liberal	Australia	8	95
	Canada	13	
	Ireland	2	
	New Zealand	2	
	Switzerland	7	
	UK	12	
	US	51	
Conservative	Austria	9	55
	Belgium	3	
	France	22	
	Germany	16	
	Netherlands	4	
	Luxembourg	1	
Scandinavian	Denmark	5	27
	Finland	5	
	Iceland	2	
	Norway	7	
	Sweden	8	
Southern	Greece	4	51
	Italy	21	
	Portugal	7	
	Spain	19	
East Asia	Japan	10	17
	South Korea	7	
East European	Czech Republic	8	24
	Hungary	3	
	Poland	8	
	Slovakia	3	
	Slovenia	2	
Not Belonging to Any Current Welfare Type	Chile	4	23
	Israel	6	
	Mexico	6	
	Turkey	7	
Total	33 nations	292 regions	292 regions

Consequently, 69 regions from six nations (Chile, Hungary, Mexico, Poland, Slovakia and Turkey) under the threshold were excluded from the cases. The other 292 regions over the threshold can remain. Another issue arises regarding how to categorise the regions to different welfare regimes. Some researchers have done pioneering works of applying the welfare regime categorisation to the sub-national local levels such as global cities (Hudson, 2012) or British and German local welfare states (Schridde, 2002). However, the 'welfare regime modelling business' at the local level is yet at its nascent stage. Inevitably, the 292 regions' welfare regime types are given in accordance with those of their mother nations as presented in Table 5-6. No regions in the first five regimes were omitted in this process.

Given relatively few studies on welfare regime types of Chile (Simon and Picazo, 2005), Mexico (Dion, 2006) and Turkey (Buğra and Keyder, 2006), they are tentatively categorised as not belonging to any of the welfare regimes as is Israel in the earlier part of this chapter.

5-4. Conclusion

This chapter is devoted to explanation and justification of selection of methods, variables, datasets and cases for the two methodological approaches. In the first approach, the pooled TSCS datasets, of which finding will be introduced in Chapter 6, is used to overcome the 'small N' problem, to capture variation of variables over both time and space, to control for time effects, and to reduce the omitted variable bias. The merits of this pooled TSCS data analysis may overcome some of the six methodological challenges discussed in the introduction of this chapter such as ecological fallacy, regional variance, small-N, outlier, heterogeneity, and historical contingency problems. However, some complications, including regional variance and historical contingency problems, still remain unaddressed.

Then this researcher uses the second multiple regression analysis with the larger regional dataset to address most of the limitations, at least partially. On the other hand, the multiple regression also has its own limitation as it focuses only on one-time dataset while the pooled TSCS analysis encompasses the over-time trend. In this context, the two approaches are complementary in producing answers to the identical research questions.

Regarding the choice of independent variables, any available datasets are selected in the statistical models under the two conditions; reliable datasets need to be available and they are theoretically regarded as health determinants (as reviewed in Chapter 2). As a result, in the first pooled TSCS design, this researcher selects a total of seven independent variables, including income per capita, income inequality, education, unemployment rate, alcohol consumption, government health spending and tax and transfer effect. In the second multiple regression model, a similar but different set of independent variables were utilised. They are income per capita, income inequality, unemployment rate, air quality and tax and transfer effect.

Six dependent variables in the pooled TSCS and multiple regression designs respectively are all health indicators. They are all from OECD datasets and chosen to examine different impact of the independent variables not only on the whole population but also on sub-groups of the population, especially depending on gender and age. In terms of case selection, this researcher conducts two rounds of correlation test to select 26 nations and 292 regions over a certain threshold of GDP per capita.

This thesis has the several contributions with regards to the abovementioned methodological choices. First, this thesis uses the OECD regional database for the first time in the international comparative health study. The larger set of the regional cases enable researchers to easily overcome the chronic "small N" (Esping-Andersen, 2007, p. 335; Ebbinghaus, 2005, p. 133) problem which has long plagued cross-national comparative researchers. The potential breakthrough enables us in this thesis to address most of the six methodological challenges.

Second, the six different dependent variables in each model can illustrate the different pathways between health determinants and health outcomes of different population sub-groups, in particular, younger and older and male and female. The detailed examination helps us to overcome the limitations of previous cross-national health studies that they tend to oversimplify the dynamic pathways by ignoring the different and often contrasting health outcomes between aged and younger generations or females and males.

Third, the case selection is processed with the statistical evidences, of which method based firmly on the theory on the curvilinear relationship between income and aggregate health (Wilkinson, 1992; Wolfson et al., 1999). This researcher selects the cases over a threshold where "for rich countries to get richer adds nothing further to their life expectancy" (Wilkinson and Pickett, 2009, p. 6). For the first time, this researcher uses the Pearson correlation tests (Adler and Parmryd, 2010) repeatedly until a threshold can be located where the correlation between income and health vanishes. The simple statistical method can resolve the thorny but often ignored problems of unjustified case selection issue (see Ebbinghaus, 2012; Kim, 2015).

The fourth and final contribution of this thesis is to incorporate the East Asian welfare regime (Aspalter, 2006; Goodman, Kwon, White, and more, 1998) in the cross-national comparative health studies. It is debatable why the East Asian welfare states as a group have rarely been included despite their noticeable health enhancement. For example, Japan has arguably the longest living people in the world for the last decades. Korea's record is also remarkable for its dramatic increase in life expectancy from 52.4 years in 1960, 16 years below the average of OECD members, to 81.3 years in 2012, well above the OECD average and even higher than those of the Netherlands (81.2) and Finland (80.7) (OECD, 2015f). Despite these achievements, the East Asian regime has long been ignored with only some recent exceptions (Chuang, Chuang, Chen, Shi, and Yang, 2012; Popham, Dibben, and Bambra, 2013). The thesis compares health outcomes of East Asian welfare regime with those of other welfare regimes.

All these potentially original methodological contributions are expected to shed brighter light on the illusive relationships between income inequality, welfare regimes and aggregate health. We will see the findings from these approaches in the next two chapters.

CHAPTER 6 POOLED TIME-SERIES CROSS-SECTIONAL ANALYSIS

6-1. Introduction

In seeking answers to the research question of the relationships between income inequality, welfare regimes and aggregate health, this chapter discusses the findings from the pooled time-series cross-national (TSCS) data analysis. As clarified in Chapter 5, this researcher uses the pooled TSCS data with eight independent variables and six dependent variables of the selected 26 nations over the period 1995~2012.

This chapter has three purposes. First, the presence of the second Scandinavian puzzle, suggested in the previous review chapters regarding Scandinavia's relatively low level of aggregate health in comparison with those of other regimes, will be examined with empirical evidence. Second, East Asian health status, long ignored in the cross-national comparative health studies, will be compared to check their better-than-expected aggregate health outcomes. Third, the pooled TSCS data analysis identifies health determinants, such as GDP per capita, that make statistically significant associations with aggregate health outcomes to analyse the dynamics behind income inequality, welfare regimes and aggregate health.

This chapter consists of the three parts. The first part describes the patterns of the six dependent variables and nine independent variables of the five key welfare regimes in 2010, the most recent timing of the time-series cross-sectional dataset with the fewest number of missing observations. The data is calculated by weighting the individual nation's population within each regime. The findings here corroborate the presence of the 'second Scandinavian puzzle'. Contrary to the theoretical expectation, the Scandinavian welfare regime reports relatively poor aggregate health despite its egalitarian welfare states. On the other hand, the East Asian welfare regime shows remarkably good health outcomes despite its relatively negative measures of health determinants such as income inequality and welfare benefits. The unexpected positive outcomes correspond with the East Asian puzzle, proposed in the previous chapters.

In the second part of this chapter, this researcher conducts a TSCS regression with AR(1) Prais-Winsten correction and panel-corrected standard errors (PCSE) for 26 industrialised nations for the 1995~2012 period. The statistical outcomes show that GDP per capita has stable and consistent effects on all the health indicators. GDP per capita maintains its strong association with aggregate health despite some claims (Wilkinson, 1992; Wilkinson, 1996) that its influence is negligible over a certain 'threshold' of high income level (i.e. the GDP threshold effect). For the other independent variables such as education, disposable income Gini or unemployment rate, their effects are limited to either old age or female group, supporting this thesis' assumptions on the age and gender threshold effects. However, market income Gini does not have any clear association with any health indicators.

In the third and last part, this researcher goes back again to the descriptive analysis of the pooled TSCS data and examine the time-series trend of the health determinant independent variables and health outcomes over the period. This section scrutinises the descriptive statistics of the independent and dependent variables over time. A further finding is that the East Asian welfare regime with its minimal rise in GDP per capita for the subject period had its average life expectancy increased by the biggest margin. On the other hand, the Scandinavian welfare regime ended up raising its life expectancy by one of the lowest margins despite its biggest economic growth. East Asian life expectancy has never been lower than the Scandinavian statistics in the subject period and the gap has been widening consistently. The findings are again in tune with the second Scandinavian puzzle and the East Asian puzzle.

6-2. Cross-national Descriptive Data Analysis

As the first step of the descriptive analysis, this chapter compares, among the five welfare regimes, the six dependent variables in 2010 as seen on Figure 6-1. (The dynamics of the variables over the 1995~2010 period will be discussed in the later 6-4 section of this chapter.) Each variable in the figure were calculated by weighting the populations of each regime's member nations in consideration of the large variations in national populations. For example, Luxembourg with its population of around half million should not be treated equally with approximately 600 times bigger United States (around 316 million). The weighted average can produce the relatively more accurate health indicators of each welfare regime (see Lynch et al., 2001, p. 196; Ebbinghaus, 2012, p. 7). Figure 6-1 presents outcomes of the six dependent variables in the smaller Figures I~VI. The definition of the each variable will be presented in Table 6-1. For better and intuitive understanding of the figures, the bars for the Scandinavian and East Asian regimes are highlighted as black and dark grey respectively with all the others light grey in the figure.

Figure 6-1. Six Health Outcomes of Five Welfare Regimes

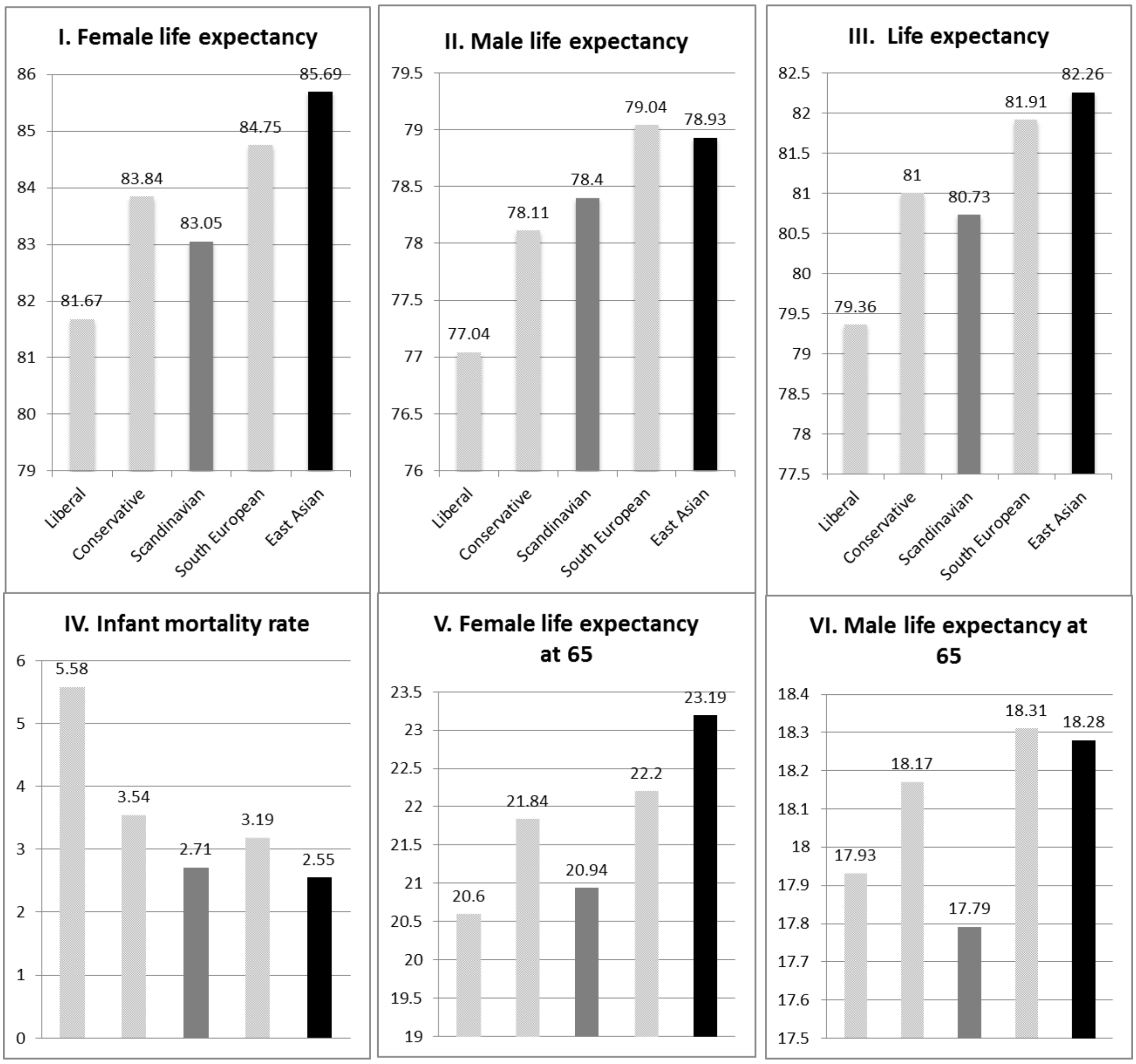


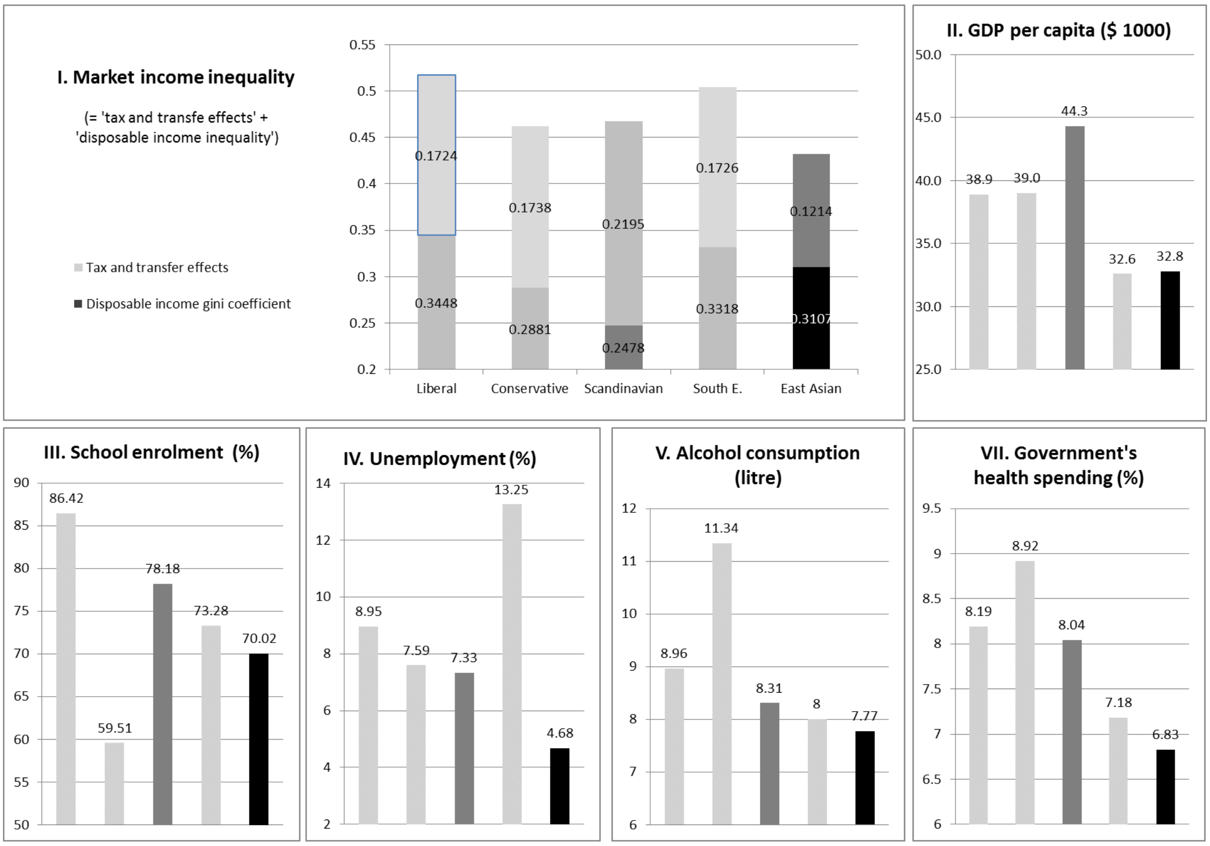
Figure I for female life expectancy shows that the black-coloured bar for the East Asian regime is the highest (85.7 years) but the dark grey-coloured one for the Scandinavian regime is shorter by 2.6 years. The other two bars for South European (84.8 years) and Conservative (83.3 years) are higher than the Scandinavian bar but shorter than the East Asian bar. The Liberal regime's bar was the shortest (81.7 years). However, when it comes to the other health indicators, the 'ranking' is rearranged. For example, in the figure II, South European males live longest (79.0 years) to be followed by East Asians

(78.9 years). Scandinavian males were again behind the two groups and ranked as the third (78.4 years).

The figures I~VI reveals the following main findings. First, the Scandinavian welfare regime's health indicators are overall worse than the general theories suggest (Wilkinson, 1992; Gravelle, 1998). The regime is not ranked the highest in any of the categories and is placed even below third place for four indicators (I, III, V and VI) out of the six. It has the worst record in terms of male life expectancy at 65 (VI). Its infant mortality is the only positive outcome ranked as the second best. It is a noteworthy contrast to the theoretical expectation on Scandinavian excellence in enhancing aggregate health presumably boosted by its equal labour market and egalitarian welfare states (Richter et al., 2012; Bambra, 2011). Even though it is generally accepted that it underperforms in narrowing health inequalities compared with other regimes, at least it is expected to show the best outcomes in terms of enhancing aggregate health (Hurrelmann, Rathmann, and Richter, 2010; Mackenbach, 2012). However, this simple descriptive statistics show radically different results. Simply, the Scandinavian welfare regime does not live up to the theoretical expectation.

On the other hand, the East Asians' health outcomes are impressive. They topped the four (I, III, IV, V) out of the six health categories and the second in the rest two categories (II, VI) of male life expectancies at birth and at 65. Even for infant mortality rate, which Scandinavian nations have generally kept the best results in most of previous studies (Chung and Muntaner, 2007; Raphael, 2013), Scandinavians are surpassed by East Asians. In no categories are East Asians are ranked below second place.

Figure 6-2. Nine Health Determinants of Five Welfare Regimes



Even in these second-placed two health categories (male life expectany at birth and at 65), the differences with the frontrunning South European males are only marginal at 0.11 and 0.03 years respectively.

The South European welfare regime has also noticeable, albeit less remarkable than its East Asian counterpart, in its health results, which are constantly as one of the best three spots in any of the categories. Conservative regimes are roughly in the middle between healthy East Asian and South European regimes and Liberal and surprisingly unhealthy Scandinavian and regimes. The Liberal welfare regime takes all the worst outcomes except for the male life expectancy, placed only the second worst.

Given this different health outcomes between the welfare regimes, the next step is to analyse the cause of the different health outcomes between the same set of welfare regimes. Figure 6-2 presents the statistics on the nine health-determinant independent variables. The definitions of the variables are presented in detail in Table 6-1 in the later section of this chapter.

Again, the East Asian regime is highlighted in black, the Scandinavian in dark grey and the others in light grey. In the first 'I. market income inequality' figure, three different variables are presented altogether as they are closely related. It has been noted that the re-distributional 'tax and transfer effects' is the deducted amount from 'the market income Gini coefficient' by 'the disposable income Gini coefficient' (Joumard, Pisu, and Bloch, 2012; Luebker, 2012). Then, the combined sum of the 'the disposable income Gini' (the lower dark-coloured part of each bar) and the 'tax and transfer effects' (the upper light-coloured part of each bar) represents the market income Gini coefficient. In the figure I, the market income inequality (the combined bar) is the lowest in the East Asian regime with its Gini coefficient at 0.4321, but the Scandinavian regime is the lowest in terms of disposable income Gini coefficient at 0.2478 with its most generous tax and transfer effects (0.2195).

Overall, the Scandinavian welfare regime has relatively impressive figures, with its rankings within the positive second place in seven categories out of the eight. It is the richest (44,300 dollars in GDP per capita). It is the second most educated (78.2 percent tertiary school enrolment rate) after the Liberal regime (86.4 percent) and its unemployment rate is also the second lowest (7.33 percent) following the East Asian (4.68 percent). Even in the rest two categories of 'alcohol consumption (8.31 litres)' and 'government health expenditure (8.04 percent)', it is placed in the third spot. The Scandinavians are not placed in the two bottom spots in any category.

In contrast, the East Asian welfare regime is the worst in two variables of re-distributional 'tax and transfer effects' and 'government spending on health.' The East Asian governments' interference in the market through re-distributional tax and transfer and public spending on health is notably weaker than even those of Liberal regime. However, East Asia has also some positive aspects with the lowest alcohol consumption and the lowest unemployment rate. The East Asian data in a respect correspond to its typical welfare regime characteristics, described as 'low unemployment and limited state welfare' (Holliday, 2005; Wilding, 2008). South Europeans also show the worst health determinants statistics in three categories of 'GDP per capita', 'market income inequality' and 'unemployment rate'.

Given these datasets on Scandinavian welfare regime's good health determinants, the finding is paradoxical in that the welfare regime reports unexpectedly poor aggregate health outcomes. In other words, good independent variables appear to be associated with poor dependent variables. As proposed in the previous systematic review chapter, these counterintuitive findings can confirm 'the second Scandinavian puzzle'. It is named such a way to follow the precedent of the first 'Scandinavian puzzle' (Bambra, 2011) regarding the regimes' surprising underperformance in narrowing health inequalities within its population compared with the other welfare regimes (Lahelma and Arber, 1994; Sacker, Worts, and McDonough, 2009). To put it simply, the first puzzle is on the health inequalities and the second, proposed in this thesis, is on the aggregate health. Combined, the Scandinavians are suffering from 'the dual

Scandinavian puzzles' in its surprising underperformance in both narrowing the health inequalities and enhancing the aggregate health.

On the other hand, the East Asians' relatively good health status, despite its less impressive health determinants statistics, poses another challenge to the conventional international comparative health theories. East Asian health outcomes are in fact dramatic contrast with the Scandinavian health outcomes in that it shows positive outcomes out of relatively negative inputs. Consequently, this additional counterintuitive finding corresponds with 'the East Asian puzzle'.

In the few studies that include the East Asian welfare regime in examining the between-regime health outcomes, researchers have been surprisingly reluctant to accept its impressive outcomes. Karim, Eikemo, and Bambra (2010, p. 45) conclude in the abstract only that "the East Asian welfare states did not have the worst health outcomes" even though East Asians have "the highest average life expectancy" (p. 51) among the compared six welfare regimes. Another study reaches the similar conclusion that "East Asian welfare states did not have worse health than most welfare states" (Chuang, Chuang, Chen, Shi, and Yang, 2012, p. e23). In their article, East Asians again have the highest average life expectancy (Chaung et al., 2012, p. e24). In the last case of such studies, Popham, Dibben, and Bambra (2013) do not provide detailed analysis on East Asian health outcomes as their research focus is on the Scandinavian health inequalities. However, female life expectancy is the highest in the 'Confucian' welfare regime in their findings.

With these backgrounds, this study is probably the first to shed light on the East Asia's best health outcomes as one of the research foci. In addition, the South European welfare regime's health records are also impressive given its lowest GDP per capita, the highest market income inequality and also the highest unemployment rate. These paradoxical findings will be discussed further in the following sections of this thesis.

6-3. Pooled TSCS Data Analysis

6-3-1. Descriptive Statistics

This researcher turns to the pooled TSCS data of the selected 26 nations over the 18 years in 1995~2012. Table 6-1 describes definitions, means, and standard deviations of the six dependent and eight independent variables in addition to their maximum and minimum values of the combined 468 country-year observations (= 26 × 18). Beside each maximum and minimum value, the related country-year is specified.

In the table, the national average income, GDP per capita, has its mean of 37.2 in thousand dollars with its standard deviation of 11.5. The figures of the 468 observations range from the lowest (South Korea's 16,580 dollars in 1995) to the highest (Luxembourg's 89,991 dollars in 2007). In the case of market income inequality, the lowest figure, Iceland statistics in 1995 (0.3088), is almost doubled in the highest Portuguese figure in 2007 (0.5629). Iceland in 1995 is again the most equal nation in terms of disposable income (0.1911) while United States report the highest disposable income inequality (0.3812) in 2008.

Table 6-1. Independent and Dependent Variables

	Definition	Mean	S.D.	Minimum	Maximum
income	GDP per capita in constant prices (in 1000 dollars)	37.2	11.5	16,580 (Korea, 1995)	89,991 (Luxembourg, 2007)
market income Gini	Gini coefficient before taxes and transfers	0.4585	0.048	0.3088 (Iceland, 1995)	0.5629 (Portugal, 2007)
disposable income Gini	Gini coefficient after taxes and transfers	0.2939	0.043	0.1911 (Iceland, 1995)	0.3812 (United States, 2008)
tax and transfer effects	tax and transfer effects	0.1646	0.047	0.0198 (Korea, 2004)	0.2732 (Ireland, 2010)
education	school enrolment rate in tertiary education (%)	61.26	18.1	7.38 (Luxembourg, 1995)	116.62 (Greece, 2012)
unemployment rate	share of the labour without work but seeking employment (%)	6.86	3.47	1.8 (Luxembourg, 2001)	25.2 (Spain, 2012)
alcohol	alcohol consumption per capita aged over 15 (in litres)	9.72	2.6	1.5 (Israel, 1995)	15.1 (France, 1995)
government health spending	government health expenditure as a percentage of GDP (%)	6.55	1.421	1.43 % (Korea, 1995)	9.9% (Netherlands, 2012)
life expectancy	average years that a person at birth can be expected to live	79.23	1.88	73.3 (Czech R., 1995)	83.2 (Japan, 2012)
female life expectancy	average years that a female at birth can be expected to live	81.91	1.93	74.8 (Greece, 1995)	86.4 (Japan, 2012)
male life expectancy	average years that a male at birth can be expected to live	76.43	2.15	69.6 (Korea, 1995)	81.6 (Iceland, 2012)
female old-age health	average years that a female at 65 can be expected to live	20.25	1.36	16.2 (Czech R., 1995)	24.0 (Japan, 2009)
male old-age health	average years that a male at 65 can be expected to live	16.74	1.33	12.7 (Czech R., 1995)	20.1 (Iceland, 2012)
infant mortality rate	deaths of children aged 0~1 per 1000 live births	4.32	1.25	0.9 (Iceland, 2011)	8.1 (Greece, 1995)

The redistribution effects through tax and transfer was the lowest in South Korea in 2004 (0.0198) and the highest in Ireland in 2010 (0.2732). The two nations are located at the extreme ends to show the dynamic between market distribution and state redistribution. In the case of South Korea, it keeps the lowest tax and transfer effects among the OECD membership but maintains one of the lowest figures of market income inequality. Its market income inequality in 2002 (0.3324) is the second lowest figure among the 468 nation-year observations. In contrast, Ireland has one of the unequal market inequality with its figure in 2010 (0.5616) being the second highest out of all the total observations. However, its highest tax and transfer effects offset the widest market income inequality. If South Korea is a radical 'generous distribution, mean redistribution' case, Ireland is the opposite 'mean distribution, generous redistribution' extreme. In the end, when it comes to the disposable income inequality which combine market distribution and state redistribution, the two cases meet in the middle in the ranking as the 14th (South Korea) and 15th (Ireland) in terms of narrow disposable income inequality out of the 26 nations in 2010. The cases of South Korea and Ireland demonstrate the dynamics between the distribution and redistribution, which justifies the use of the two indicators as discrete independent variables in this thesis.

The school enrolment in tertiary education is the lowest in Luxembourg in 1995 with only 7.38 percent and the highest at 116.62 percent in Greece in 2012. The low Luxembourg figures seem to be related to its lack of a university until the creation of "Université du Luxembourg" in 2003 (European Commission, n.d.). The Greek figure could exceed 100 percent because it is defined as "the ratio of total enrolment, regardless of age, to the population of the age group that officially corresponds to the level of education shown" (World Bank, 2010, p. 109). The Unemployment rate has its average at 7.25%. Luxembourg in 2001 has the lowest unemployment rate (1.8%) and it is the highest in Spain in 2012 (25.2%).

In the case of life expectancy, the average of the available 466 observations is 79.23 years with a standard deviation at 1.88. It is the shortest in Czech Republic in 1995 (73.3) and the longest in Japan in 2012 at 83.2. In terms of female life expectancy at birth and at 65, Japan in 2012 tops the list with its figure at 86.4 years and 24.0 years respectively. On the other hand, another island nation, Iceland has its male population live the longest in 2012 as its life expectancy at birth (81.6 years) and at 65 (20.1 years) indicate. The infant mortality rate has its mean of 4.32 deaths out of 1000 live births with its standard deviation at 1.25. It is the lowest at 0.9 in 2011 in Iceland, but the highest at 8.1 in 1995 Greece.

6-3-2. Pooled TSCS Regression

Given the statistics of the 14 variables, the regression models of this chapter are as follows.

Model 1)

Y_{it} = β_1 + β_2 log(GDP per capita$_{it-5}$) + **β_3 (disposable income Gini$_{it-5}$)** + β_4 (school enrolment rate$_{it-5}$) + β_5 (unemployment rate$_{it-5}$) + β_6 (alcohol$_{it-5}$) + β_7 (government health spending$_{it-5}$) + α_i + λ_{t-5} + ε_{it}

Y denotes the health indicators (male, female and combined life expectancy at birth, infant mortality rate, female and male life expectancy at 65) for country i at time t. On the right-hand side of the equation, each indicator represents an independent variable for country i at time t-5. Then, α_i is the country specific effect and λ_{t-5} is a time-specific effect. The last ε_{it} is the error term. The five-year time lag between independent (t-5) and dependent variables (t) is designed in the models.

In the case of studies linking income inequality and health outcomes, the majority of the studies use the contemporaneous dataset for both dependent and independent variables. Zheng (2012) finds that 59 out of 63 studies on the impact of income inequality on aggregate-level health assume the instantaneous effects in their research designs. However, critics raise issues regarding the presumption (Subramanian and Kawachi, 2004; Zheng, 2012). The studies focusing on the possible lagged effects conclude that the effects are "the strongest from five years and up to 15 years later" (Subramanian and Kawachi, 2004), "peak at 7 years" (Zheng, 2012), are present "up to 15 years" (Blakely et al., 2000) or are "both contemporaneous and lagged" (Macinko, Shi, and Starfield, 2004). This thesis follows the precedent study's (Ram, 2006) design of assuming a five-year lag between independent and dependent variables to attenuate "the simultaneity problem" (p. 781). Consequently, the dependent variable in 2005 in our dataset will be analysed in connection, not with the contemporary 2005 independent variables, but with the past 2000 variables. The adjusted setting inevitably reduces the number of observations from 468 (= 26 nations × 18 years) to 338 (= 26 nations × 13 years). The period for independent and dependent variables are respectively reduced to 1995~2007 and 2000~2012.

Then another Model 2 is added. The only difference between Models 1 and 2 is that this second model divides the 'disposable income Gini' variable (in bold in Model 1) into the two variables of 'market income Gini' and 'tax and transfer effects' to see their separate effects on aggregate health.

Model 2)

$Y_{it} = \beta_1 + \beta_2$ (GDP per capita$_{it-5}$) + $\boldsymbol{\beta_3}$ **(market income Gini$_{it-5}$) +** $\boldsymbol{\beta_4}$ **(tax and transfer effects$_{it-5}$)** + β_5 (school enorolment$_{it-5}$) + β_6 (unemployment rate$_{it-5}$) + β_7 (alcohol$_{it-5}$) + β_8 (government health spending$_{it-5}$) + $\alpha_i + \lambda_{t-5} + \varepsilon_{it}$

As seen in the equation, the pooled TSCS data combines the spatial and temporal dimension. This pooled data has several advantages to analyse the dynamics between variables across units over times as discussed in Chapter 4. However, the complex nature of the 'stacked' dataset also poses five technical challenges because the dataset's intrinsic characteristics violate the basic assumptions prerequisite for ordinary least square (OLS) regression, the most common form of the regressions. The five complications due to the violations or other factors are summarised as below (see Hicks, 1994, pp. 171~174; Podesta, 2002, pp. 9~11).

1) Unit or period effect: there can be unspecified unit or period effects underlying in the pooled TSCS dataset. The nations and periods are assumed to be homogenous for OLS regression, but the heterogeneous traits can exist between nations or periods. For example, in this thesis, some health-determining variables such as dietary habit or environmental quality can be different between nations but they could not be included as independent variables due to data unavailability. The omitted variables may inflate or deflate the estimates of the coefficients of some variables included in the statistical model.

2) Autocorrelation: errors of one unit in one time tend to be dependent on that of the previous time. It is because UK's population data in 2010, for example, cannot be

expected to be independent from its value in 2009, and, in turn, the 2009 data not free from the 2008 one. In fact, the temporally successive values tend to show a certain trend, resulting in an autocorrelation.

3) Contemporaneous autocorrelation: errors tend to be correlated across nations. While the autocorrelation above is related to time, this is a spatial correlation. For example, "We would not expect errors for Belgium to lack some resemblance to those for the Netherlands or errors for Canada and the United States to be altogether independent" (Hicks, 1994, p. 174).

4) Heteroscedasticity: Nations with high values may have a higher error variance. For example, South Korea and Mexico with higher GDP show larger variances in manufacturing outputs than those of Singapore or Greece with lower GDP (Dougherty, 2011, pp. 283~284). The general assumption for OLS regression is the homoscedasticity.

5) Causal heterogeneity: "errors may tend to be non-random across spatial and/or temporal units because parameters are… are heterogeneous across subsets of units" (Hicks, 1994, p. 172). If we take one example from this thesis, the estimated influence of income inequality on aggregate health, expressed in terms of the slope coefficient, can vary over time and space. However, pooled TSCS analysis may average out the differences or dynamics (Kittel, 1999) posing a 'pooling dilemma' or the "to pool or not to pool" problem (Podesta, 2002, p. 27).

In case these complications are not found in the pooled TSCS data, the solution can be easier and we can run the pooled OLS regression. In the OLS regression, all the observations from the different time periods are pooled as a single sample. The simplest model has been in fact preferred as it is "often as good or better than more complicated ones" (Beck and Katz, 2004, p. 27). However, the OLS regression is likely to be inefficient or biased because the regression does not recognise the structure of N nations of T years but treats each of NT (NT=N X T) nation-year unit as an independent observation (Stimson, 1985, p. 921). Therefore, this researcher needs to test the presence of the violations to avoid the inefficiency or bias in our analysis.

The outcomes of the tests are illustrated in Table 6-2. At first, this researcher can test the presence of the unit or period effects using pFtest by comparing the fixed effects model and the pooled OLS models (Croissant, Millo, and others, 2008). With its alternative hypothesis supporting the significant unit or period effects, its p-value nears to zero when life expectancy is a dependent variable. It means that the presence of unit or period effects is statistically significant. As the next step, this researcher tests if the unit or period effects are random or fixed by using the Hausman test (Hausman, 1978; Wooldridge, 2001, pp. 288~291). With the null hypothesis that there is no correlation between independent variables and unobserved effects, the test rejects the hypothesis with its p-value nearing to zero. Under the presence of the correlation, fixed effects model, rather than random effects, model produce the consistent estimator. In fact, econometricians warn that random model is not desirable for non-random sample such as the OECD member nations (Wooldridge, 2008; Dougherty, 2011). Therefore, the fixed effects model is tentatively selected for this thesis.

Second, this researcher also checks possible serial correlation of the residuals. The serial correlation occurs when the values of the error term (ε_{it}) do not have independent distributions over time (Dougherty, 2011). This serial correlation is common because the value of one error term in one observation at t is often likely to be related to its value at t-1. The presence of the serial correlation should be tested and adjusted because it can cause erroneous estimation of standard errors (Dougherty, 2011, pp. 429~433). The Breusch-Godfrey/Wooldridge test for serial correlation in panel models (Croissant, Millo, and others, 2008) is conducted for the model with life expectancy as a dependent variable. The presence of the autocorrelation is statistically significant.

Third, the Pesaran CD test (Pesaran, 2004; Croissant et al., 2008) is conducted to examine the presence of contemporaneous correlation between nations. The test rejects the null hypothesis of no correlation between units. Fourth, the Breusch-Pagan test (Breusch and Pagan, 1979; Hothorn, Zeileis, Millo, and Mitchell, 2010) shows that the presence of the heteroscedasticity is statistically significant. For the fifth and last causal heterogeneity problem, it is assumed that the potential estimation problem is not serious because the relatively homogenous OECD member states (see Plümper, Troeger, and Manow, 2005: 353) are selected over the latest period of 1995~2012 and do not test its presence.

Given all the complications as presented, this researcher cannot use the pooled OLS regression and need to use other statistical methods to address all these 'noises'. Among several models suggested by researchers to control for them, the Parks-Kmenta model' proposed by Parks (1967) and elaborated by Kmenta (1971) has been the most popular among social scientists for decades (Hicks, 1994). However, Beck and Katz (1995) contend that the model has a critical problem of its standard errors leading to extreme overconfidence and often underestimating variability by 50 percent or more. They propose an alternative estimator of the standard errors called 'panel corrected standard error' (PCSE), which they claim works well in their Monte Carlo simulation.

Table 6-2. Pooled TSCS Data Analysis Outcomes

Complications	Test name	R command	Statistical outcomes	Meaning
Fixed effect 1	pFtest	pFtest	F=134.3***	Significant fixed effects present
Fixed effect 2	Hausman	phtest	χ2=76.1***	Fixed effects, rather than random effects, present
Autocorrelation	Breusch-Gorfrey/Wooldridge	pbgtest	χ2=238.5***	Significant autocorrelation present
Contemporaneous correlation	Pesaran CD test	pcdtest	z=5.25***	Cross-sectional dependence present
Heteroskedasticity	Breusch-Pagan	bptest	BP=267 ***	Heteroskedasticity present

The PCSE estimator since then has been highly influential in the following cross-national panel studies and is installed in many statistical software packages as a standard procedure (Chen, Lin, and Reed, 2010). However, as Beck and Katz (2004) clarify, the PCSE is not a 'panacea' and they correct the only two TSCS problems: 3) contemporaneous correlation and 4) heteroskedasticity. For the other two problems of 1) unit or period effects and 2) autocorrelation, they suggest additional solutions (Beck, 2001; Beck and Katz, 2004).

First, unit or period dummy variables can address the problem of the unspecified country and year effects. Second, to control for autocorrelation, a lagged dependent variable (LDV) can be added as one of independent variables. The two techniques together with PCSE has been so largely used among social scientists that it is called as "canonical" (Shalev, 2007, p. 285) or "de factor Beck/Katz standard" (Plümper, Troeger, and Manow, 2005, p. 327).

Table 6-3. Complications and Solutions in TSCS Data Analysis

Complications of OLS estimations	Solutions suggested by Beck & Katz (1996, 2004)	Solutions in this thesis, suggested by Plümper et al. (2005)
Heteroskedasticity/ Contemporaneous Correlation	Panel Corrected Standard Error (PCSE)	Panel Corrected Standard Error (PCSE)
Autocorrelation	Lagged Dependent Variable	Two-step Prais-Winsten feasible generalized least squares (FGLS)
Unit or Period Effects	Unit or Period Dummy Variables	Not using the dummy variables and taking the resulting risk

In this thesis, this researcher uses the PCSE to account for 3) contemporaneous correlation and 4) heteroscedasticity. However, this researcher uses neither country or year dummy variables nor LDV to address the problems of 1) unit or period effects or 2) autocorrelation. Table 6-3 summarises the four complications of pooled TSCS data analysis and the corresponding solutions suggested by the Beck/Katz group and another set of solutions adopted in this thesis, mainly suggested by Plümper et al. (2005).

Regarding choice of the solutions, at first, there are three reasons not to use the dummy variables. First, critics point out that the technique can absorb too much cross-sectional or over-time variance (Huber and Stephens, 2001; Plümper et al., 2005). For example, in the case of this thesis's model, GDP per capita rises throughout the period in most of the nations, synchronising with the similarly increasing life expectancy, which indicates the strong correlation between the two variables. However, if we add the period dummy, it can erase or significantly absorb the GDP effects, distorting the statistical outcomes.

Second, unit dummies prevents us from analyse the effect of time-invariant exogenous variables because the model is designed to eliminate the effects of the unchanging independent variables (Dougherty, 2011, p. 518).

Third, the technique's exclusive reliance on changes in levels, not on levels themselves, can be misleading. For example, in this thesis, the alcohol consumption level is included as an independent variable because it is supposed to be related to health. However, what the fixed effects model, using the unit dummy variables, is focusing on is not 'the alcohol consumption level' but 'the changes in the levels over time'. Notably, "if a theory predicts level effects, one should not include unit dummies" (Plümper et al., 2005, p. 334). Given the potential three problems, this thesis forgoes the dummy variables and takes the risk of getting estimators biased due to the underlying unit or period effects.

In the case of autocorrelation, this thesis also takes a different method from the Beck/Katz standard that proposes the inclusion of a lagged dependent variable (e.g. Y_{t-1} or Y_{t-5}). There are two reasons. First, the use of LDV is 'atheoretical' (Huber and Stephens, 2001; Wawro, 2002). As long as the current dependent variable figure is not influenced by the previous data, the insertion of an LDV among an array of independent variables seems "more an afterthought than a reasoned model specification decision firmly grounded in theory" (Wawro, 2002, p. 47). For example, in this thesis, we can assume that the life expectancy in 2010 is related by the 2009 data, but can NOT expect the present data is 'influenced' by the data one year ago. Second, LDV can falsely dominate a regression (Achen, 2000; Plümper et al., 2005). Achen (2000, p. 1) notes that "when an autoregression term is put in "as a control," it often acquires a large, statistically significant coefficient and improves the fit dramatically, while many or all

Table 6-4. Pooled Time-Series Cross-Sectional Regression (Model **1**)

Independent Variables	Dependent Life expectancy	Variables Female life expectancy	Male life expectancy	Female life expectancy at 65	Male life expectancy at 65	Infant mortality rate
ln(GDP per capita)	3.086***	2.547***	4.193***	1.771***	1.928***	-0.000037*
	[0.575]	[0.644]	[0.702]	[0.375]	[0.415]	[0.000016]
Disposable income Gini	2.507	– 2.573	2.601	3.213*	5.013**	5.562*
	[2.044]	[2.062]	[0.021]	[1.459]	[1.577]	[1.605]
School enrolment ratio	0.023***	0.021**	0.025***	0.015**	0.017***	– 0.017*
	[0.006]	[0.006]	[0.006]	[0.005]	[0.004]	[0.007]
Unemployment	-0.029	– 0.019	– 0.025	– 0.003	– 0.034	– 0.005
	[0.026]	[0.030]	[0.027]	[0.020]	[0.020]	[0.026]
Alcohol consumption	– 0.098**	– 0.029	– 0.142***	– 0.012	– 0.090***	0.018
	[0.036]	[0.030]	[0.042]	[0.024]	[0.027]	[0.032]
Gov't health spending	0.136*	0.088	0.154*	0.081	0.138*	– 0.026*
	[0.065]	[0.076]	[0.074]	[0.055]	[0.055]	[0.080]
Intercept	45.74***	55.14***	31.55***	-0.088	– 5.193	6.900***
	[6.289]	[7.044]	[7.427]	[4.058]	[4.376]	[1.460]
Observations	307	307	307	308	308	301
R^2	0.998	0.998	0.997	0.973	0.973	0.750
Wald statistics (6)	47.38***	42.20***	75.86***	33.81***	73.54***	91.19**

p<0.001***,<0.01**,<0.05*.

of the remaining substantive coefficients collapse to implausibly small and insignificant values."

Given the limitations of the LDV insertion method, this thesis does not use it and instead uses the Prais-Winsten transformation to eliminate serial correlation of errors following the suggestion by Plümper et al. (2005). The Prais-Winsten correction regression can be suited to this thesis' model as "the dependent variable is trend-ridden" and, this author believes, as far as this thesis' model is concerned, that "the explanatory variables can explain the trend" (Plümper et al., 2005, p. 349).

Table 6-4 shows the outcome of the panel regression model with AR(1) Prais-Winston correction and panel corrected standard error (PCSE). The software package is panelAR in R (Kashin, 2014). In the table, the first column shows the six independent variables and the top row the six different dependent variables. Then each row presents the association between an independent variable and all the dependent variables. The figures in the brackets indicate the panel-corrected standard errors.

R^2 indicates "the proportion of the total sum of squares explained by the regression line" or "goodness of fit" (Dougherty, 2011, pp. 103~107). In the table, R^2 figures are all over 0.97, except for the model with infant mortality rate, indicating the apparently high explained sum of squares. These high R^2 figures are mainly due to the similar over-time trend between the health indicators and other independent variables including GDP per capita and educational attainment as demonstrated in Figure 6-3 and 6-5. In time-series regression models, when one dependent variable has a set of some explanatory variables growing simultaneously, the R-squared tends to have a very high figure over 0.9 (Yunker, 2000, p. 110). Wald statistics is a test for the overall model significance in consideration of both heteroscedasticity and serial correlation (Wooldridge, 2008, p. 812) and the statistics shows the statistical significance of the models.

Out of all the independent variables, the first GDP per capita turns out to be statistically significantly associated with all the health indicators. For example, its coefficient 3.086 indicates in the level-log regression that one percent rise in GDP per capita is related to a 0.031 ($\simeq$ 0.03086) year increase in life expectancy at birth. Similarly, a one percent rise in average income is also expected to bring an increase in female life expectancy by 0.025 year, in male life expectancy by 0.042 year, elderly female life expectancy by 0.018 year, and elderly male life expectancy by 0.019 year. Based on the statistics, the average income may influence aggregate health throughout life course for both male and female groups. However, its influences are relatively weaker for female population.

School enrolment ratio is another consistent health determinant with its association statistically significant for all the health indicators. A one per cent point increase in school enrolment rate is statistically significantly associated with a 0.021 year rise in female life expectancy, a 0.025 year in male life expectancy, a 0.23 year in combined life expectancy, a 0.015 year in female life expectancy at 65, a 0.017 year in old-age male life expectancy, and a 0.017 drop in infant death case. Unlike the GDP per capita, this school enrolment rate looks to have a relatively similar effect on females and males.

Other independent variables have their impact statistically significant only for a portion of population. In the interesting case of alcohol consumption, its impacts are significant only for half the population, namely males. For example, a one litre increase in terms of pure alcohol consumption (not in alcoholic beverage) is associated with a 0.14 year drop in male life expectancy and a 0.09 year decrease in old male life expectancy. Alcohol consumption does not have statistically significant effects on any female health indicator or infant mortality rate.

Government health spending shows a similar pattern. It has statistically significant associations with only male health indicators in addition to infant mortality rate. A one per cent point increase in government spending on health as percentage of a nation's GDP might lead to a 0.15 year rise in life expectancy, a 0.14 year in old male life expectancy and a 0.26 cases drop in infant mortality rate. Its impacts are not statistically significant for female health.

The effects of disposable income Gini are challenging. The associations between the income distribution index and old-aged life expectancy are counterintuitive because the rise in income inequality turns out to raise both female and male life expectancy at 65 in statistically significant ways. One possible hypothesis is that health inequalities in association with income inequality might account for the paradoxical findings. People aged over 65, from presumably more favourable socioeconomic backgrounds than those who died before 65, might benefit further from the widening income inequality and consequently live longer under wider income inequality. However, the hypothesis is still debatable because some researchers note wide income inequality not only hurts the poor but also the rich (Subramanian and Kawachi, 2006). Then the counterintuitive findings in this thesis may need further elaboration. However, disposable income inequality also has a statistically significant negative association with infant mortality rate. A 0.01 point rise in disposable income Gini is related to 5.56 infant death cases. Disposable income inequality has a theoretically expected relation with infant mortality rate but its relationships with old-age health indicators are counterintuitive.

This researcher also runs the 'Model 2' with the second 'disposable income Gini' variable split into the two separate variables of 'market income Gini' and 'tax and transfer effects'. The outcomes are almost identical with the 'Model 1' with GDP per capita and school enrolment ratio both having relatively consistent effects on all the health indicators. Alcohol consumption and public health spending have statistically significant associations only with male health indicators. It is not surprising in that the market income Gini is the sum of 'disposable income Gini' and 'tax and transfer effects'.

Table 6-5. Pooled Time-Series Cross-Sectional Regression (Model **2**)

Independent Variables	Dependent Variables: Life expectancy	Female life expectancy	Male life expectancy	Female life expectancy at 65	Male life expectancy at 65	Infant mortality rate
ln(GDP per capita)	3.067***	2.489***	4.286***	1.842***	2.030***	– 0.000034*
	[0.554]	[0.593]	[0.713]	[0.374]	[0.418]	[0.000016]
Market income Gini	0.021	– 0.038	0.022	0.024	0.042**	4.699
	[0.020]	[0.022]	[0.021]	[0.015]	[0.016]	[2.455]
Tax & Transfer Effects	– 0.038	0.014	– 0.033	– 0.046*	– 0.059**	– 7.215**
	[0.023]	[0.022]	[0.023]	[0.018]	[0.018]	[2.393]
School enrolment ratio	0.022***	0.021***	0.025***	0.015**	0.018***	– 0.016*
	[0.006]	[0.006]	[0.006]	[0.005]	[0.004]	[0.007]
Unemployment rate	– 0.038	– 0.009	– 0.018	0.009	– 0.021	0.009
	[0.023]	[0.030]	[0.027]	[0.023]	[0.020]	[0.025]
Alcohol consumption	– 0.104**	– 0.037	– 0.141**	-0.016	– 0.092***	0.013
	[0.035]	[0.028]	[0.042]	[0.024]	[0.027]	[0.032]
Gov't health spending	0.151*	0.101	0.158*	0.100	0.141*	– 0.234**
	[0.064]	[0.074]	[0.074]	[0.055]	[0.055]	[0.087]
Intercept	46.29***	60.00***	30.81***	-0.369	– 5.912	7.260***
	[6.097]	[6.424]	[7.521]	[4.030]	[4.403]	[1.550]
Observations	307	307	307	308	308	301
R^2	0.998	0.998	0.997	0.974	0.973	0.753
Wald statistics (6)	55.24***	50.17***	78.39***	41.01***	78.93***	100.16**

pvalue<0.001***,<0.01**,<0.05*

Therefore, there is not, and should not be, any major differences between the outcomes between the two models. However, we can observe the influence of the new independent variables of market income inequality and tax and transfer effects.

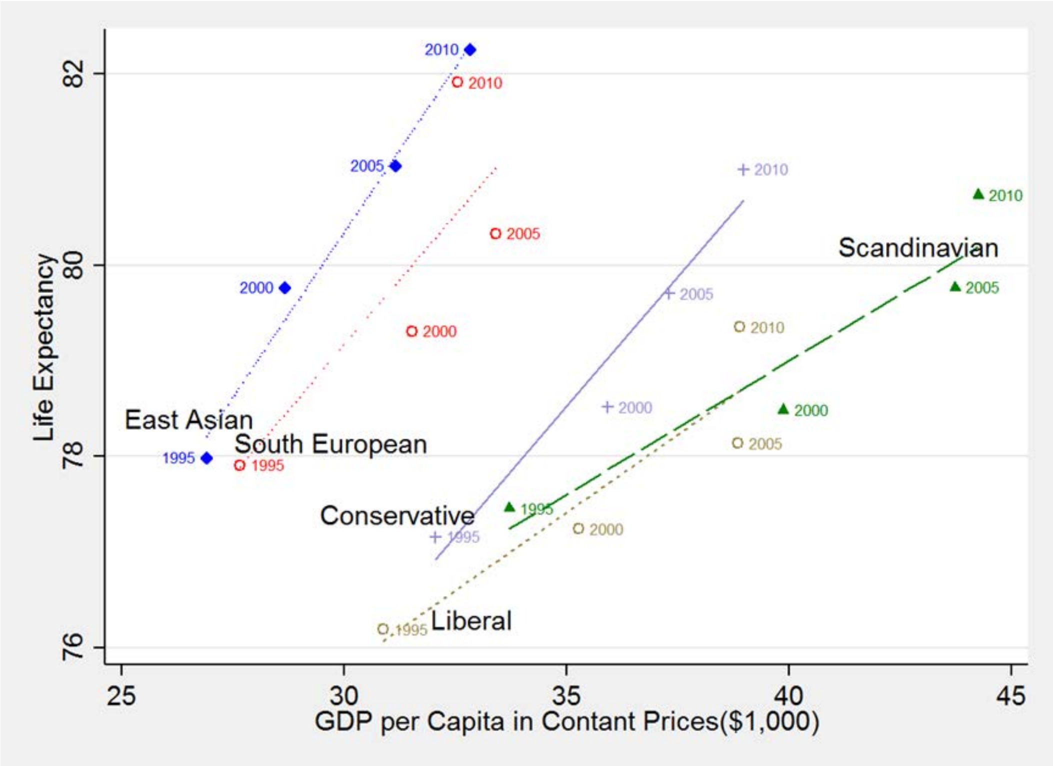
First, market income inequality does not have any statistically significant association with any of the health indicators except for a counterintuitive relationship with old-age life expectancy. According to the finding, the wider the market income inequality, the longer the old male population lives. This is consistent with the finding on the counterintuitive relation between disposable income Gini and old-age health in Model 1. Second, the tax and transfer effect has identical effects with disposable income Gini, having statistically significant and negative effects on old-age health outcomes but a positive effect on infant mortality rate. It is noteworthy that infant mortality rate is regarded as "the most sensitive to political and welfare state conditions" (Chung and Muntaner, 2007, p. 331). On the other hand, the old-age health indicators have counterintuitive associations with all the income inequality indicators.

6-4. Cross-Regime Descriptive Analysis over Time

In the pooled TSCS analysis in the previous section, it is found that GDP per capita and education have constant influences on all the health indicators. We turn to the dynamics of the TSCS data over the time periods. The data is again weighted by the population of each nation to calculate the weighted average of each welfare regime just like it has been done in the earlier section of this chapter. The descriptive analysis is expected to shed light on the dynamics between the health determinants and health outcomes over the period. As the first step, this researcher examines the changes in life expectancy and GDP per capita of the five welfare regimes as shown in Figure 6-3.

The figure shows paradoxically contrasting two relationships between GDP per capita and life expectancy. At first, all the regimes have their weighted average life expectancies increased over the 15 years along with their increases in GDP per capita.

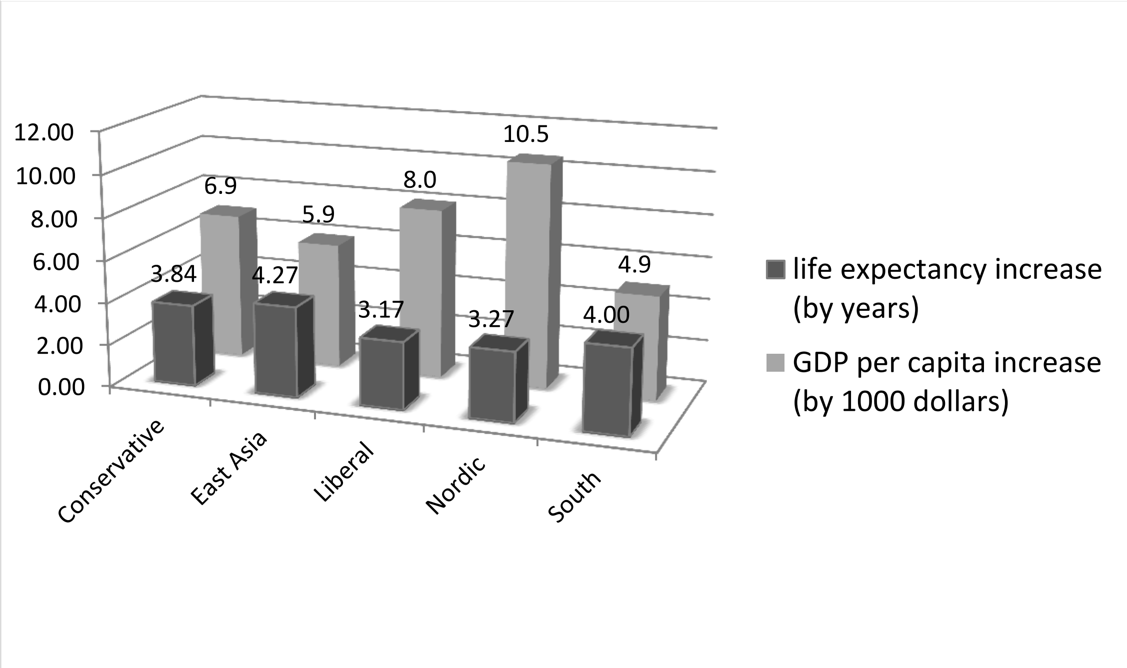
Figure 6-3. GDP and Health



There is no exception apart from the South European Regime during the financial turmoil in 2005~2010. In this 'within regime' perspective, life expectancy might be hypothesised as closely related with average income.

The second and contrasting trend is about the between-regime differences in each time points. In 2010, for example, the poorest two regimes, East Asian and South European models, with their average income hovering just above 30,000 dollars and far fewer than those of the other regimes, live longest by around three years more than the richest Scandinavian regime with the second shortest life expectancy. In this perspective, it can be also paradoxically hypothesised that "the poorer, the healthier". In the other three timings of 1995, 2000 and 2005, the hypothesis can be observed with only few exceptions.

Figure 6-4. Increases in Income and Life Expectancy



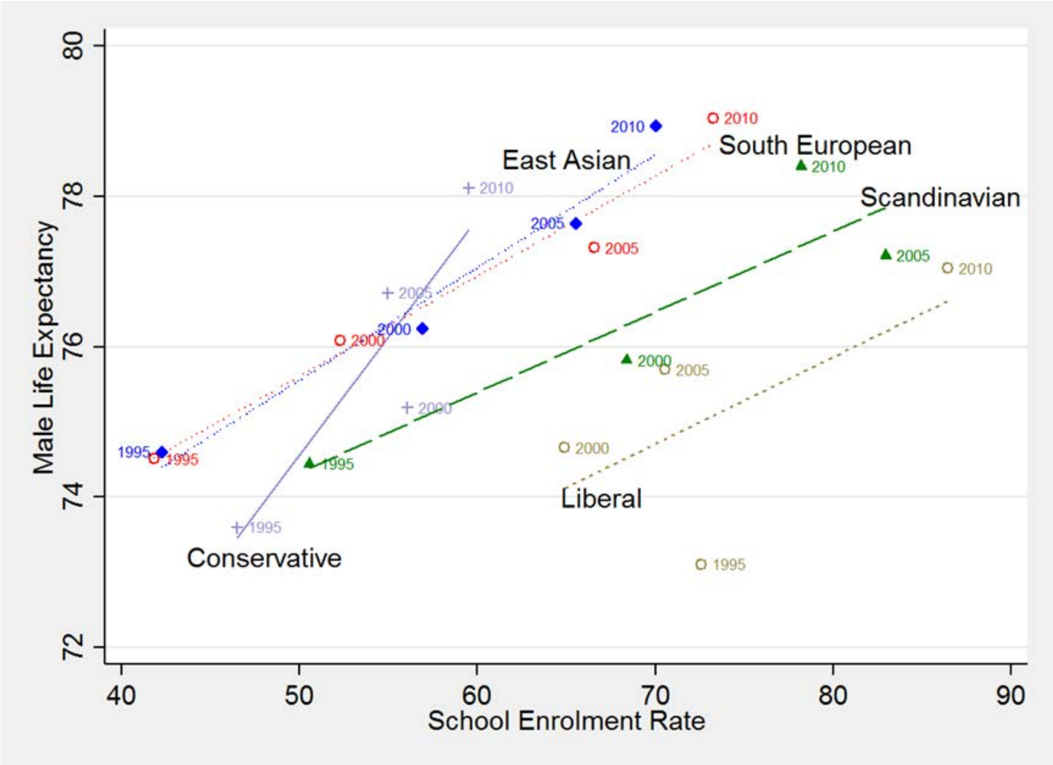
It is noteworthy that 'the richer, the healthier' hypothesis is supported in the 'within regime' time-series observations, while the counterintuitive 'the poorer, the healthier' hypothesis can be seen in the contemporaneous cross-sectional observations. This is a typical case of Simpson's paradox (Simpson, 1951). The paradox occurs when the direction of association between variables X and Y reverses after pooling over a covariate Z (Dong, 2011, p. 852). In this case, X is income, Y is life expectancy and Z is welfare regime. The paradox in this context visualises the second Scandinavian puzzle and the East Asian puzzle because the former rich live shorter and the poor latter live longer in the figure. In other words, Figure 6-3 vividly illustrates an anti-theoretical aspect of the two puzzles. If we accept the two puzzles, the counterintuitive hypothesis of 'the poorer, the healthier' might be acknowledged.

It is even more challenging if we go deeper and observe the dynamic trends of life expectancy and GDP per capita in the five regimes for the 1995~2010 period. Back in Figure 6-3, the gradient for East Asian trend is steeper than those of any other regimes, meaning that it had increased then already the longest life expectancies by the biggest margin for the last 15 years. Figure 6-4 illustrates the differences between the regimes over the period.

The Scandinavian welfare regime increased their GDP per capita by the biggest margin (10,542 dollars), but their life expectancies rose only by 3.27 years, only 0.1 year more than the worst Liberal regime's average increase. In East Asia, the average income rose by the second smallest margin (5,924 dollars), but their life expectancy rose by 4.27 years, the biggest increase. It can be partly due to South Korea's rapid 7.1 year increase in life expectancy on the back of its fast growing economy, but the longest-living Japanese also raised their longevity by 3.3 years, higher than the average increase of either Liberal or Scandinavian welfare regime (OECD, 2015f).

Theoretically, East Asians with their already highest life expectancy, are expected to show relatively marginal increases in life expectancy over the period, but what we are observing is simply to the contrary. The East Asian welfare regime's sharp increase in longevity widens the gap with those of the other regimes. This deepening gap poses challenges for the conventional comparative health studies. South Europeans were also 'efficient' in enhancing their aggregate health with relatively less increase in average income. If the trend persists, the gap between East Asia and the relative laggard Scandinavian might be widened even further in the future. It follows that both the second Scandinavian puzzle and the East Asian puzzle may persist.

Figure 6-5. Education and Male Life Expectancy



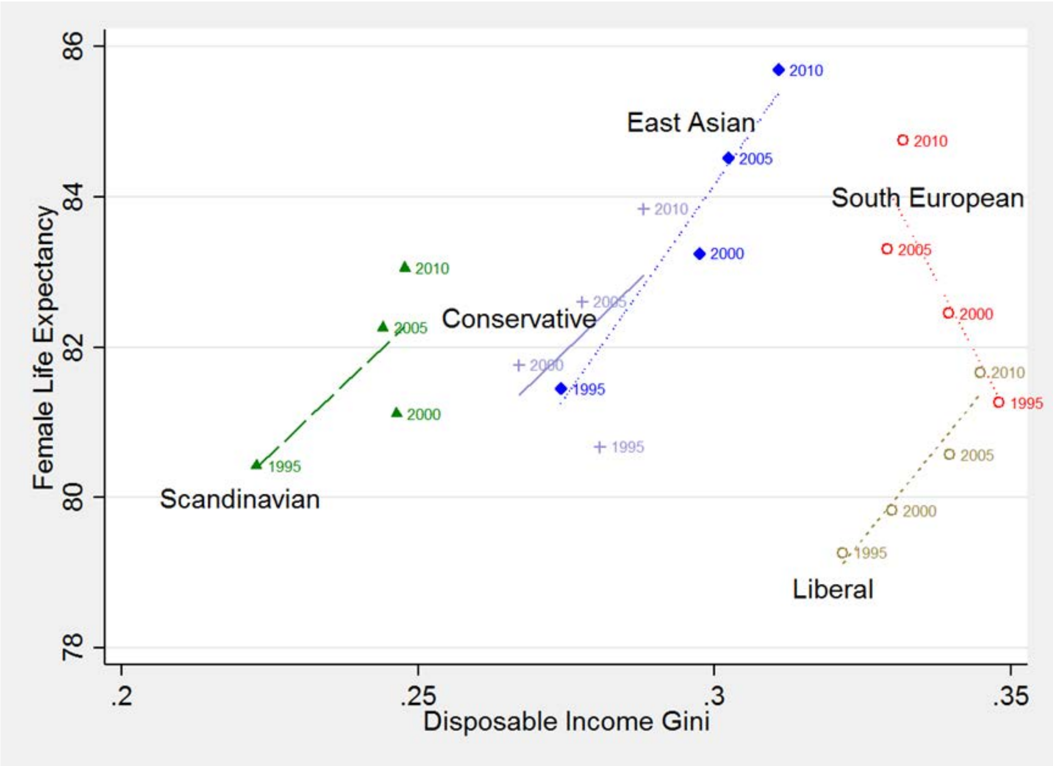
However, the finding in the previous panel-corrected data analysis gives a clue what is NOT an answer for the puzzles, because we find that income is a strong and consistent health-enhancing variable. The Scandinavian welfare regime does not enhance its aggregate health as much as that of the East Asian welfare regime 'despite' its biggest growth in income. Given that, we may be able to assume that the other variables, certainly unfavourable to Scandinavian health, are strong and persistent enough to overshadow the economic benefits favourable to its aggregate health. Then a question arises: what are the other variables?

The pooled TSCS analysis shows that education, operationalised as the tertiary enrolment rate, is a consistent predictor of health with its impact slightly stronger for male population than female counterparts. Table 6-5 demonstrates the relationship between male life expectancy and the education variable over the 15 years. Again, we can observe Simpson's paradox along the trajectories of the five welfare regimes.

At first, all the welfare regimes increased the average enrolment rates over the period with only three exceptions: the Liberal regime in 95~00, the Scandinavian welfare regime in 05~10, and the Conservative regime in 00~05. The drop in the Scandinavian model seems to be related to the introduction of tuition fee for students from outside the EU, which began in Denmark in 2006 and Sweden in 2011 (Välimaa, 2015). According to OECD (2012, p. 4), "there has been a significant decline in the enrolment rate of students" from out of Europe in the major Scandinavian nations.

In the meantime, all the welfare regimes have their average male life expectancy increased constantly. Consequently, we might be able to assume that the more educational attainment is related to the longer life expectancy as long as this 'within regime' perspective is concerned. However, 'between regime' examination leads to the opposite finding. For example, in 2010, the less educated East Asian and South European welfare regimes have longer life expectancies than the more educated Scandinavian and Liberal welfare regimes. This counterintuitive trend can be also applied to the other timings of 1995, 2000 and 2005. According to this 'between regime' comparison, we may reach a paradoxical 'the less educated, the healthier' inference. The Simpson's paradox, observed in the previous relationship between income and life expectancy, reemerges in the relationship between education and male life expectancy. This paradox is observed for female life expectancy as well.

Figure 6-6. Disposable Income Gini and Female Health



Again, as we have found in the pooled TSCS regression that education is the consistent and statistically significant health determinants, we assume that there must be 'other variables' that counteract the favorable impacts of education for Scandinavian health and offset the maleficent effects of income and education for East Asian health.

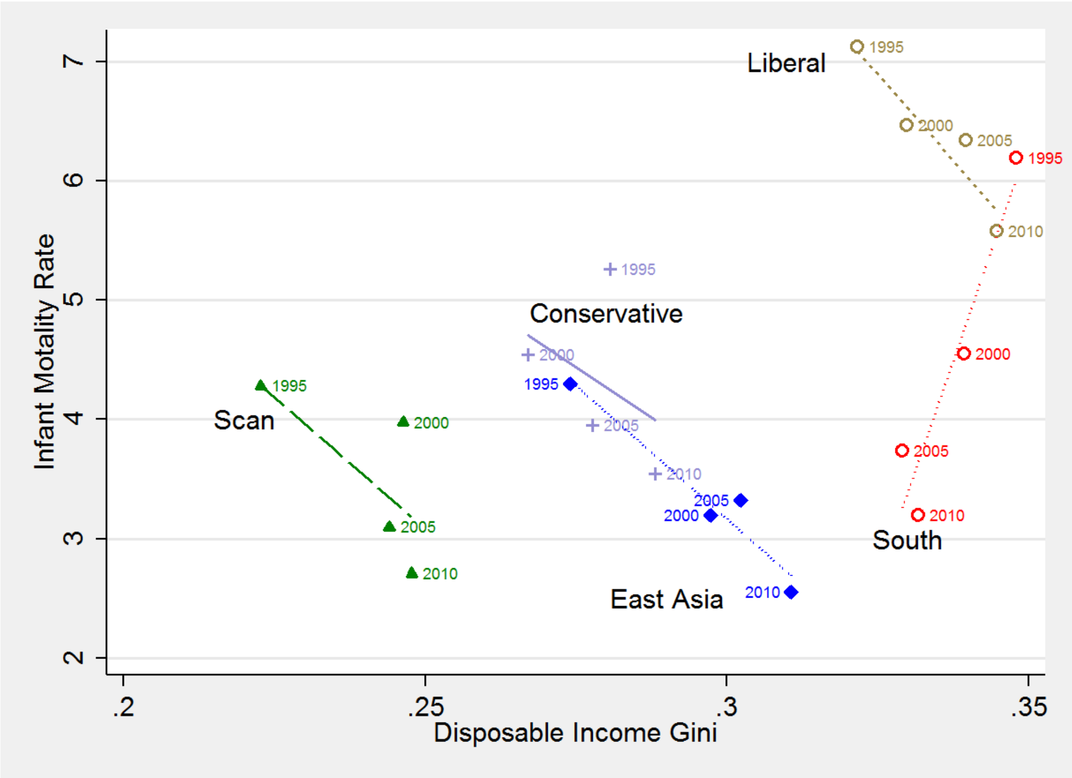
Income inequality can be one of 'other variables'. However, the pooled TSCS regression in this chapter cannot find any statistical significance between any of the three income inequality indicators and any of the aggregate health indicators except for infant mortality rate. Considering some claims that female health tends to be more sensitive to income inequality aggravation compared with males (e.g. Nowatzki, 2012; Torre and Myrskylä, 2014), this researcher at first examines the trends between female life expectancy and disposable income Gini in Figure 6-6.

In the figure, the Gini indicator generally aggravate for all the welfare regimes except for South European welfare regime, which narrowed the disposable income gap over the 15 years. Scandinavians come as the second in terms of income inequality aggravation, widened by 0.025 point. However, it is still apparent that the Scandinavian welfare regime is undisputedly the most egalitarian in terms of the disposable income inequality despite its recent widening trend. Then theoretically their life expectancy should still have risen by the biggest margin in the meantime because they may endure the least maleficent effects of inequalities (Wilkinson, 1992; Wilkinson and Pickett, 2009).

On the other hand, the East Asian welfare regime has its disposable income Gini widened by far the biggest margin (0.037) among the welfare regimes. In addition, the East Asian regime, with constantly higher disposable income inequality than that of Scandinavian model, should have suffered more from the maleficent effects from income inequality. However, they have in fact succeeded in raising their longevity by the biggest margin. Again, it is obvious that disposable income inequality does not provide any clue to the second Scandinavian puzzle and the East Asian puzzle.

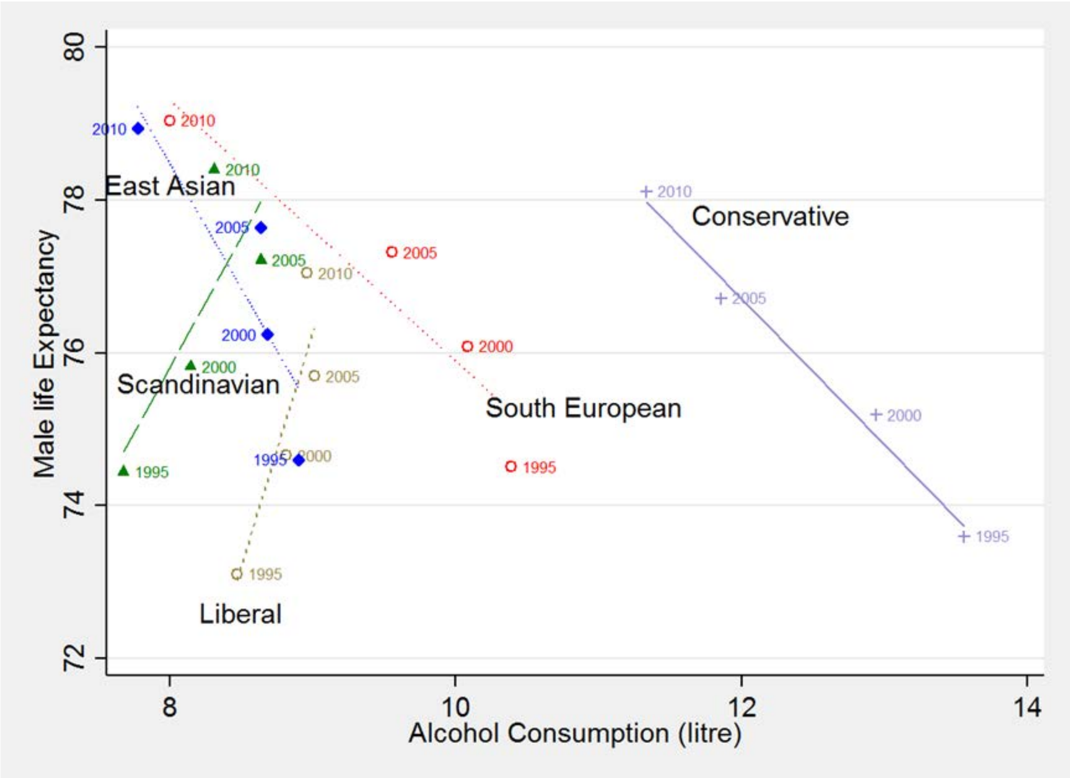
In Figure 6-6, the 'within regime' trend is counterintuitive in that the growing disposable income distribution accompanies the rising female life expectancy. The 'between regime' comparison does not signal any visual relationship. This simple analysis of the figure at least partly explains why disposable income does not show any statistically significant relationship with any health indicators except for infant mortality in the previous pooled TSCS regression.

Figure 6-7. Disposable Income Gini and Infant Mortality



On the other hand, Figure 6-7 illustrates why disposable income inequality has a statistical association with infant mortality rate in the pooled TSCS regression. Again, the 'within regime' perspective generally shows the decreasing infant mortality in step with the widening disposable income inequality. However, the 'between regime' aspect demonstrates the negative effects of wide market income inequality on aggregate health. The Scandinavian welfare regime with the visually fairer disposable income distribution enjoys the lowest infant mortality rate, while the most unequal Liberal regime has the worst health indicators. These cross-regime differences may mirror the statistically significant impact of disposable income inequality.

Figure 6-8. Alcohol Consumption and Male Life Expectancy



Similarly, this researcher examines the relationships between the nine independent variables and the six dependent variables one by one after presenting the relationships like Figures 6-3~7. Some figures are noteworthy. We have a factor in explaining the East Asian excellence: alcohol consumption, as seen in Figure 6-8. Japan and South Korea has reduced their alcohol consumption to the lowest level for the 15 years, while the Scandinavians have drunk gradually more and their alcohol consumption is higher than that of South Europeans in 2010. People in Liberal and Scandinavian welfare regimes increase their alcohol consumption during the subject period. The alcohol consumption in the Conservative regime is visually higher than those of any other regimes. The health-related behavior may give a tiny, but not vital, clue in analyzing both the East Asian and Scandinavian puzzles.

6-5. Conclusion

This chapter analyses the pooled TSCS dataset to seek answers to the relationships between income inequality, welfare regimes and aggregate health. As the first step, the cross-regime descriptive analysis with the population-weighted average of each regime shows that the Scandinavian welfare regime with its relatively good health determinants statistics paradoxically do not show the best health outcomes. On the other hand, the East Asian welfare regime with its relatively poor health determinants succeeds in enhancing its aggregate health more than those of any other welfare regime. The findings corroborate presence of the second Scandinavian puzzle and the East Asian puzzle.

Second, this researcher uses the panel-corrected standard error (PCSE) and AR(1) Prais-Winsten correction to correct bias and inefficiency in estimating coefficients of variables in the pooled TSCS dataset. After analysing the 338 nation-year observations of 26 nations over 13 years, it is found that income and education is generally consistent health determinants. This finding refutes the proposed 'GDP threshold effect' because the income does not change its pattern of influence on aggregate health in wealthy societies. However, the effects of income inequality, alcohol consumption and public health spending are limited only part of population, depending on gender and age. The inconsistent effects on the subgroup of population again support the hypothesised threshold effects of health determinants over age and gender.

Finally, the third descriptive TSCS statistics show that income, income inequality, education, public health spending are variables expected to enhance Scandinavian aggregate health, but they don't provide any clues on the second Scandinavian puzzle or the East Asian puzzle except for the low alcohol consumption in East Asia. The regional datasets in the next chapter might offer an answer to the puzzles.

CHAPTER 7 MULTIPLE REGRESSION WITH REGIONAL DATASET

7-1. Introduction

In order to address the research questions on the relationship between income inequality, welfare regimes and aggregate health, this researcher has conducted the pooled TSCS regression with panel-corrected standard errors in the previous chapter. We find that GDP per capita is a constant and significant variable affecting the aggregate health over time, but income inequality is not, except for the relationship between disposable income inequality and infant mortality rate. In this chapter, this researcher turns to the multiple regression analysis with a bigger number of cases from the OECD regional dataset to further examine the research questions. As discussed in Chapter 5, we have the data covering five health determinant variables and six health indicators as dependent variables for the selected 292 regions from the 34 OECD member states. The regions are categorised as one of the six welfare regimes or 'not belonging to any of the regimes'.

This multiple regression shows that Scandinavian welfare regime underperforms in meeting theoretical expectations in all of the six health indicators, while East Asian welfare regime outperforms the general expectation in most of the health outcomes. The statistical outcome with more observations, again, confirms the presence of the second Scandinavian and the East Asian puzzles. In addition, GDP per capita and three income inequality measures (disposable income Gini, market income Gini and tax and transfer effects) have a consistent and statistically significant relationship with most of the health indicators. Again, there is no GDP threshold effect observed.

7-2. Descriptive Statistics

Table 7-1 shows the definitions, averages, standard deviations, missing values and the minimum and maximum statistics of the 13 variables. For each maximum and minimum value, the name of related region is written within the brackets. For example, the regional income, which means 'GDP per capita in current prices (in 1000) dollars', has its average at 35.92 with its standard deviation at 15.54. It has six missing values. Poland's southern region of Lesser Poland is the poorest among the cases with its GDP capita at 17.34 while the richest is the US capital, the District of Columbia which has roughly ten times more income of 171.48. Its extreme wealth can be visualised on top left corner of the first boxplot in Figure 7-1. As seen, the average GDP per capita of the 'Liberal' regions is the highest, but the South Europe regions are the poorest on average. The boxplots also show the large variance within each regime. For example, some regions in the richest Liberal regime group are poorer than the average of the poorest South European welfare regime's GDP per capita. On the other hand, some rich regions of the poorest South European groups have their incomes higher than the average of the richest 'Liberal' welfare regime group's GDP per capita. It should be cautioned that the boxplot does not incorporate the weighted value of the region's population and only depicts the rough means and variations of the regional indicators.

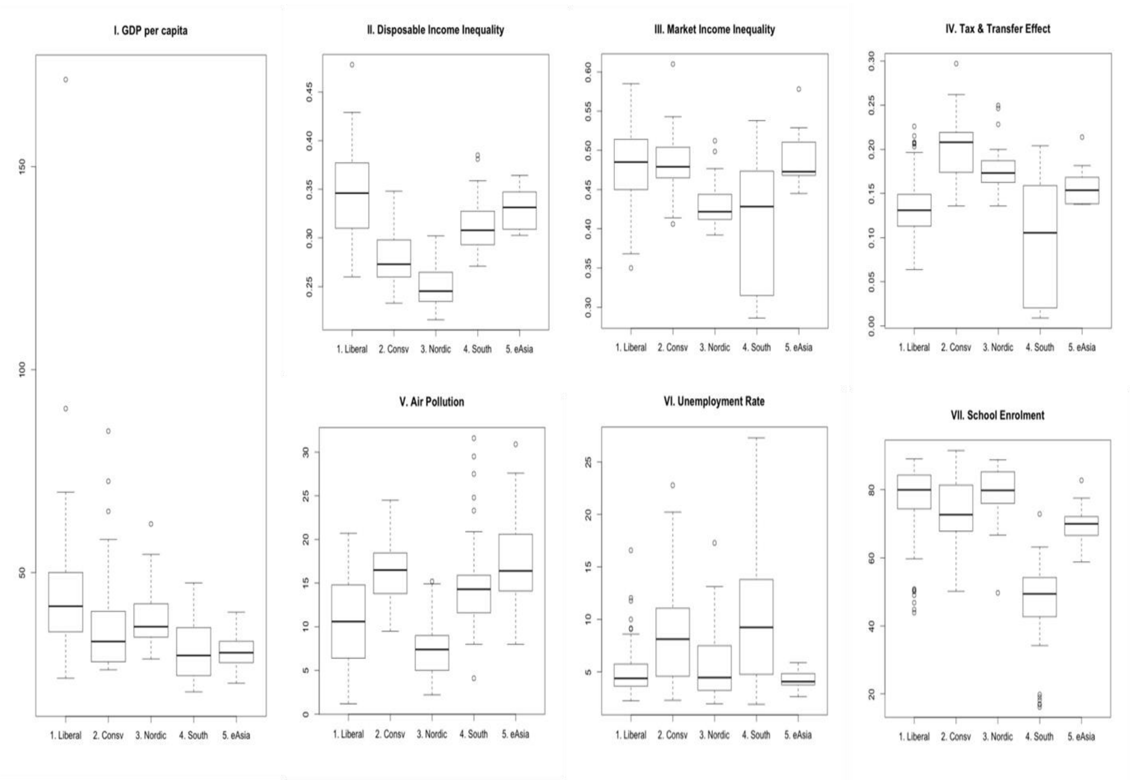
Table 7-1. Independent and Dependent Variables

N= 292	Definition	Mean	S.D.	Missing values	Minimum	Maximum
income	GDP per capita in current prices (in 1000 dollars)	35.92	15.54	6	17.34 (PL: Lesser Poland)	171.48 (US: District of Columbia)
market income Gini	Gini coefficient before taxes and transfers	0.4567	0.059	32	0.2860 (IT: Veneto)	0.61 (FR: Corsica)
disposable income Gini	Gini coefficient after taxes and transfers	0.3167	0.057	23	0.2160 (NO: Hedmark and Oppland)	0.5220 (CL: Santiago Metropolitan)
tax and transfer effects	difference between market Gini and disposable Gini	0.1383	0.059	32	0.009 (IT: Emilia-Romagna)	0.2970 (DE: Bremen)
education	share of labour force with at least secondary education (%)	70	17.1	12	16.1 (PT: Madeira)	95.7 (CZ: Prague)
unemployment rate	the ratio between unemployed persons and labour force	7.539	5.18	0	1.2 (ME: Quintana Roo)	27.31 (IT: Calabria)
air	pm 2.5: fine particulate matter in micrograms per cubic metres	13.76	5.95	0	1.2 (CA: Nunavut)	31.6 (IT: Lombardy)
female life expectancy	average years that a female at birth can be expected to live	82.85	2.295	10	76.8 (TR:Eastern Marmara - South)	87 (ES: Navarra)
male life expectancy	average years that a male at birth can be expected to live	77.34	2.54	14	70.1 (PL: Lodzkie)	81.2 (FI: Åland)
combined life expectancy	average years that a person at birth can be expected to live	80.22	2.304	7	74 (ME: Tabasco)	84 (ES: Navarra)
infant mortality rate	deaths of children aged 0~1 per 1000 live births	4.541	2.23	20	1.7 (AT: Carinthia)	16.2 (CA: Northwest Territories)
child death rate	deaths per 100 000 population of same age (0~14) group	0.425	0.22	13	0.1 (ES: Melilla)	2.1 (CA: Nunavut)
old-age death rate	deaths per 10 000 population of those aged over 65	4.4	0.52	0	2.98 (KR: Jeju)	6.35 (PT: Azores)

* A grey cell indicates a negative extreme

In addition, this researcher omits the boxplots of East European Welfare Regime in Figures 7-1~2 because their health indicators are so visibly worse than those of the other welfare regimes that, if we include East European boxplot, we cannot see the subtle differences between the other five regimes. For the purpose of the visual simplicity, this researcher only presents the five welfare regime's boxplots as our research focus is mainly on the Scandinavian and East Asian welfare regimes. The market income Gini's average stands at 0.4567 with its standard deviation of 0.059. It is the lowest in the northern Italian region of Veneto. Italy's figures in terms of market income fairness are remarkable. Of the top 20 regions in their narrowest market income inequality, 18 are Italian regions. The two exceptions are the 17th Australian Capital Territory (0.35) and the 20th East Switzerland region (0.368). Of the Scandinavian regions, Sweden's Central Norrland was the top on the list ranked only as 30th in the list (0.392). The Scandinavian regions do not have the lowest market income Gini indexes but the Southern European regions do. In fact, the third boxplot on Figure 7-1 illustrates that the South European welfare regime has such a wide variance that some regions have the extremely low data and others have relatively high statistics. The figure also reveals that not all Scandinavian regions have the lower market income inequality data than those of the 'Liberal' regimes. Some Scandinavian regions have even higher market income inequality than the average of Liberal regime's market income inequality indexes. It is in general because urban areas in any welfare regime tend to have wider Gini coefficients than rural areas. For example, Stockholm (0.4380), Helsinki (0.4497), Oslo (0.4430), Copenhangen (0.4449) in Scandinavian nations have the wider market income inequality than some regions in Liberal regime such as New Hampshire (0.4380) and Wyoming (0.4340) in U.S. or Australian Northern Territory (0.3730).

Figure 7-1. Health Determinants of Five Welfare Regimes



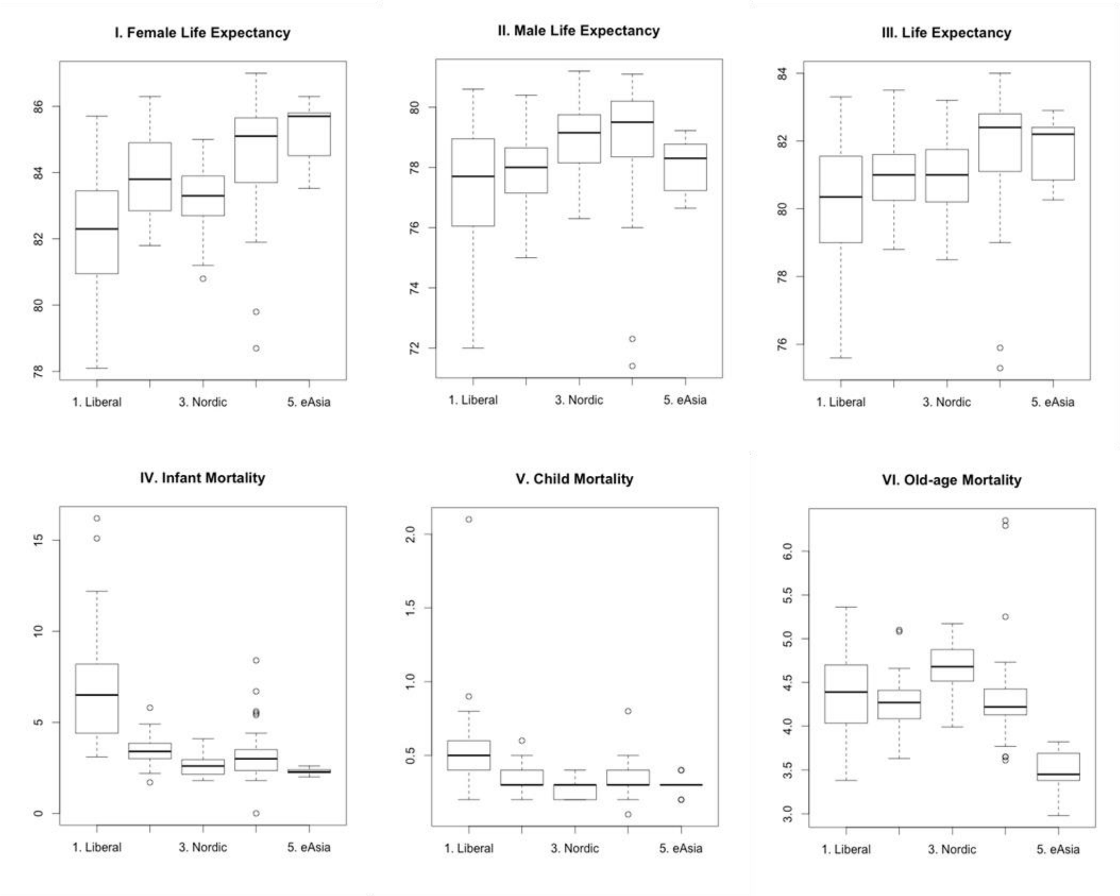
On the other end, France's Mediterranean island of Corsica has the widest market income inequality (0.61) followed by the US District of Columbia (0.585) and Japan's Hokkaido (0.5781).

Italian regions have the lowest figures in terms of the re-distributional taxes and transfers with its 13 regions among the 20 least generous ones. Italy's northern Emilia-Romagna region has the lowest statistics of 0.009, far lower than the 292-region average of 0.1383 with the standard deviation at 0.059. The most generous region with the largest re-distributional effects is German region of Bremen (0.2970). The French and German regions are relatively active in their redistribution with five German and ten French regions are included in the top 20 list. Notably, the two nations are categorised as the Conservative regime, not the Scandinavian regime. Among the top 20 regions, only three Scandinavian regions are placed (Finland's Eastern and Northern Finland, Southern Finland and Western Finland). The rest two regions are, interestingly, 'North East England' (14^{th}) and 'North West England' (19^{th}), parts of the Liberal regime. Again, the finding is contrary to the expectation that Scandinavian regions would report the highest re-distributional outcomes, but it was the Conservative regions that do so.

It is only when this researcher combines the effects of market income and redistribution that Scandinavian regions begin to emerge as the most equal. In terms of the disposable income, Norway's landlocked region of Hedmark and Oppland is the most equal region with its figure at 0.216. Out of the 20 most equal regions, twelve are from the Scandinavian regime (five from Norway, three from Sweden, four from Denmark). On the second boxplot of Figure 7-1, we can see the North European regions have their disposable income Gini generally lower than those of the other regimes. Chile's capital Santiago Metropolitan area is the most unequal (0.522) followed by Mexico's Campeche (0.521), Tabasco (0.494) and US District of Columbia (0.478). Of the worst regions, six were located in Mexico. The average of the disposable income Gini is 0.3167 and standard deviation is 0.057.

The Czech capital of Prague has the most educated labour force, as its share of labour force with at least secondary education stands at 95.7 percent. East European regions are rich with educated workforce since they take 10 spots out of the top 20 regions (six from Czech Republic, three from Slovakia and one from Poland). Portugal's seven regions (Madeira, Central Portugal, Azores and others) are the least schooled with their statistics ranged 16.1~19.8 percent. The rate average is 70.0 percent with its standard deviation at 17.1.

Figure 7-2. Health Outcomes of Five Welfare Regimes



Air quality is the best in Canada’s northernmost region of Nunavut with only 1.2 µg/m^3 in pm2.5, but the worst in Lombardy, the Italy’s industrial hub (31.6 µg/m^3).

Among the six health indicators, female life expectancy is the highest at 87 in Navarra, a northern territory of Spain, followed by other Spanish neighbours of Madrid (86.6) and Castile and León (86.5). Out of the top 20 regions, six are Spanish regions (La Rioja, Cantabria, Basque Country), six are French regions (Ile de France, Pays de la Loire, Corsica, Aquitaine, Midi-Pyrénées and Rhône-Alpes), four Japanese regions (Hokuriku, Chugoku, Kyushu-Okinawa, Hokkaido) and four Italian regions (Marche, Bolzano-Bozen, Trento, Umbria).

It is noticeable that all these Spanish regions are located in the northern territories of the nation, while most of the French regions are the nation's southern areas except for the capital area. Their geographical proximity is particularly impressive and needs further analysis in future studies because their geographical or cultural factors may contribute to their longevity. On the other hand, women living in Turkey's Mediterranean west region live the shortest life (76.8 years). Of the bottom 20, seven are from Turkey, two are from Mexico, and one each from Portugal, the Czech Republic and Slovakia. It is also remarkable that all the rest eight are US states (West Virginia, Alabama, Oklahoma, Mississippi, Kentucky, Louisiana, Tennessee and Arkansas). As seen in the first boxplot in Figure 7-2, the Liberal regime has particularly large variance in female life expectancy compared with the other welfare regimes. The average female life expectancy is 82.94 with its standard deviation of 2.18.

Finland's small island of Åland has the longest living males (81.2). All the Japanese regions, located in the best female longevity list, vanish from this male top 20 group while the five Swiss regions enter into the shortlist. The other regions are from eight Italian regions (Marche, Bolzano-Bozen, Tuscany, Trento, Emilia-Romagna, Umbria, Basilicata and Apulia), four from Spain (Navarra, Madrid, La Rioja and Castile-La Mancha), one each from Australia (capital territory) and France (Île-de-France). The average is 77.48 years and standard deviation is 2.41. The second boxplot in Figure 7-2 shows that East Asian male's life expectancy is relatively lower than those of their Scandinavian and South European counterparts.

With both genders combined, Spain's Navarra is the best in terms of longevity (84). Out of the top 20 regions, 12 are impressively from the South European regimes (Navarra, Madrid, La Rioja, Castile and León, Castile-La Mancha from Spain, Marche, Bolzano-

Bozen, Trento, Tuscany, Umbria, Emilia-Romagna, Basilicata from Italy). In addition, three regions (Île-de-France, Midi-Pyrénées, Rhône-Alpes) are from the Conservative Regime, three (Ticino, Lake Geneva Region and Central Switzerland) from the Liberal Regime, only one (Åland) from Scandinavia and one from East Asia (Hokuriku). Among the bottom 20, three are from the Liberal regime (Mississippi, West Virginia, Alabama), two are from the South European regime (Potugal's Azores and Madeira) and six from the East European regime (Poland's Lodzkie, West Pomerania, Silesia, Czech Northwest region, Slovakia's West Slovakia and Central Slovakia). Of the six regimes, South Europeans have the wide variation, having regions widespread from top to bottom together with those from the Liberal regime.

The infant and child mortality rates are unexpectedly high in Canada's northern territories. Its northernmost area of Nunavut has the highest child death rate at 2.1 deaths per 100,000 population aged 0~14. Canada's Northwest Territories also has the highest infant mortality rate of 16.2 deaths per 1000 live births. However, the remote area's population is only around 40,000 respectively and one or two cases of death at the target age can heavily influence the outcomes. Apart from the extreme cases, nine regions (five Mexican and four Turkish) have the single-digit child death rate with all the other regions have its rate below the 1.0 level. Spain's tiny autonomous territory in Africa, Melilla has the lowest child death rate at 0.1 and 23 other regions have the death rate at 0.2. They are from various welfare regimes including one Liberal, three Conservative, nine Scandinavian, seven South European and two East Asian ones.

The infant mortality rate is the lowest in Austria's Carinthia with 1.7 deadly cases per 1,000 births. The lowest 20 list again include the regions from all the welfare regimes except for the Liberal regime. They are seven Scandinavian regions, seven South

European regions, two East Asian regions, and even three East European regions. Twelve US states were included among the 20 worst infant mortality rate list along with six Mexican regions and the two abovementioned Canadian regions (Nunavut and Northwest Territories). The Liberal regime has the noticeably high infant mortality rate in general but its variance is also wide.

When it comes to the old-age mortality rates, East Asians stand out with all of the Korea's seven regions are included in the top 20 list along with Japan's four regions (Southern-Kanto, Kansai, Hokkaido and Toukai regions). South Korea's southernmost Island of Jeju has the lowest deaths rate of 2.98 per 10,000 aged over 65. On the other hand, the regions with the highest old-age death rates were mostly from East Europe (13 regions) plus two Turkish regions and, noticeably, three South European regions (Portuguese Alentejo, Madeira and Azores). Norway's Hedmark Oppland also has the 19th highest old-age mortality rate. The region also has one of the lowest child mortality rates as well. The central Norwegian region may have the paradoxical reputation as the best to live for babies and the worst for older people. The special case indicates the regional status can change sensitively depending on what health indicators are used.

As the preliminary step before moving on to the multiple regression, this researcher briefly conducts the Dunnett test to compare the health outcomes between the focus five welfare regimes as done for Figures 7-1~2. The boxplots in figure 7-2 could not take into account the weighted average value since it regards all the regions as an equivalent observation despite their various populations. With the Dunnett test (Cardinal and Aitken, 2013), this researcher can examine the statistically meaningful differences of their population-weighted means. The Dunnett test sets one particular group as the reference and compares it with each of the other groups (Lindman, 2011).

This researcher sets the Scandinavian welfare regime as the reference to be compared with the other regimes because it is the focus regime of this thesis and in addition its health outcomes were roughly in the middle among the five welfare regimes, a good location as the barometer. Each welfare regime's health indicator is weighted by its regions' populations in the analysis. It would be misleading to treat both the biggest California (with 37 million population) and the smallest Åland in Finland (with fewer than 30,000 residents) as the equivalent single dot. The Dunnett model outcomes are presented in Table 7-2.

The Dunnett model tests the mean differences between the welfare regimes. The first column indicates the mean differences between individual regimes with the Scandinavian regime are hypothesised to be zero. In each of the other column, the six health indicators under the comparison are presented respectively. For example, the first line of the statistical outcomes shows the estimated mean differences between Liberal and Scandinavian regimes. The figures in the square brackets are standard errors.

As noted under the table, the asterisk sign shows the significance level. In terms of the life expectancy, only the East Asian welfare regime has the significantly higher longevity than the Scandinavian regime. The other differences in life expectancy are not statistically significant. In female life expectancy, both South European and East Asian women are living statistically significantly longer than the Scandinavian females, but there are not any statistically significant difference observed in male life expectancy between welfare regimes. In the younger age group statistics, only the child mortality rate in the Liberal regime is statistically significantly higher than that of the Scandinavians.

Among the health indicators, the old-age mortality rate is the most puzzling. All the welfare regime figures turn out to be significantly lower than that of the Scandinavian statistics. In other words, the old-age mortality is statistically significantly the worst in the Scandinavian welfare regime. This paradoxical finding is in fact in parallel with the finding in the previous chapter where the Scandinavian life expectancy at 65 is the worst for male and the second worst for female (see Figure 6-1 in Chapter 6). On the other hand, the East Asian regime has its three health indicators significantly higher than those of the Scandinavian regime. Again, we observe another finding directly related to the second Scandinavian puzzle and East Asian puzzle, in particular for female and old-age health outcomes.

7-3. Multiple Regression

There have been several assumptions for the multiple regression. The first one is the normality assumption that the residual (the difference between the predicted and the observed values) follows a normal distribution (e.g. Dougherty, 2011, p. 114; Hair, Black, Babin, Anderson, and Tatham, 2005, pp. 103~107). This researcher conducts Shapiro-Wilk test (Shapiro and Wilk, 1965) to check the possible violation in every regression model.

Table 7-2. Mean Differences between Welfare Regimes: Dunnett Contrasts

Linear Hypotheses	Health Indicators Life Expec-tancy	Female L.E.	Male L.E.	Infant mortality rate	child mortality rate	L.E. at 65
Liberal - Scandinavian	-0.7169 [0.5499]	-0.9777 [0.5953]	-1.3973 [0.6961]	3.9469 [0.5576]	0.232*** [0.0373]	– 0.296* [0.1303]
Conservative-Scandinavian	0.3018 [0.5712]	0.7652 [0.6150]	-0.4962 [0.7191]	0.7816 [0.5790]	0.061 [0.0387]	– 0.4443** [0.1353]
South European-Scandinavian	1.246 [0.5857]	1.6785* [0.6284]	0.5292 [0.7399]	0.4272 [0.5937]	0.043 [0.0397]	– 0.476*** [0.1387]
East Asian-Scandinavian	1.3031* [0.5791]	2.3753*** [0.6156]	-0.2053 [0.7198]	-0.405 [0.5936]	0.016 [0.0387]	– 1.2639*** [0.1355]

p<0.001***,<0.01**,<0.5*

. The second is the homoscedasticity assumption: "dependent variable(s) exhibit equal levels of variance across the range of independent variable(s)" (Hair, Black, Babin, Anderson, and Tatham, 2005, p. 107). When the assumption is violated, the complication is called heteroscedasticity, which can result in inefficient OLS estimators and at the same time erroneous estimators of the standard errors of the regression coefficients (Dougherty, 2011, p. 283). This researcher conducts Breusch-Pagan test (Breusch and Pagan, 1979) to detect the heteroscedasticity.

In addition, This researcher also tests the possible multicollinearity between the independent variables. If the correlation between the independent variables are high, it will increase the population variance of the distributions of their coefficients, leading to the greater risk of obtaining erratic estimates of the coefficients (Dougherty, 2011, p.

165). In response to this possible noise, this researcher conducts the variance influence factors (VIF) test (Hair et al., 2005, p.251).

The last and forth is the linearity issue which assumes that "nonlinear effects will not be represented in the correlation value between variables" (Hair, Black, Babin, Anderson, and Tatham, 2005, p. 109). In other words, a dependent variable is basically expected to have a linear, not curvilinear, relationship with an independent variable.

To detect any possible nonlinear relationship, this researcher examines the scatterplot matrix of the variables. It can be noticed that the relationship between GDP per capita and three life expectancy indicators is curvilinear. It is in line with the findings on the non-linear relationship between income and life expectancy with every increase in income resulting in gradually smaller rise in life expectancy (Preston, 1975; Gravelle, 1998).

For the better fit of the regression line, the curvilinear relationship is linearised by log-transforming the independent variable of GDP per capita (Dougherty, 2011, pp. 197~206). Except for the GDP per capita, the other variables are not transformed as any other transformations do not seem to lead to better fit. Then this regression would be the 'level-log model' meaning "a regression model where the dependent variable is in level form and (at least some of) the independent variables are in logarithmic form" (Wooldridge, 2008, p. 845).

In consequence, the regression model is tentatively presented as the following equation. The selection process of independent and dependent variables are discussed in Chapter 2 and 5.

Model I)

$Y = \beta_1 + \beta_2$ (log(GDP per capita)) + β_3 (disposable income Gini) + β_4 (school enrolment rate) + β_5 (unemployment rate) + β_6 (air) + ε

In addition, dummy variables are added for each of the six welfare regimes and a group of 'other' regions which are not categorised as any of the six types. The Scandinavian regime is a "reference category" (Dougherty, 2011, p. 231) again as in the previous Dunnett test. Model II can be formulated as below.

Model II)

$Y = \beta_1 + \beta_2$ (log(GDP per capita)) + β_3 (disposable income Gini) + β_4 (school enrolment rate) + β_5 (unemployment rate) + β_6 (air quality index) + δ_1(Liberal) + δ_2 (Conservative) + δ_3 (South European) + δ_4 (East Asian) + δ_5 (East European) + δ_6 (Others) + ε

Again, like the previous chapter, this researcher splits the effects of disposable income inequality into market income inequality and tax and transfer effects (i.e. redistribution effect). Model III shows the level-log regression model with one more independent variable.

Model III)

$Y = \beta_1 + \beta_2$ (log(GDP per capita)) + **β_3 (market income Gini) + β_4 (tax and transfer effects)** + β_5 (school enrolment rate) + β_6 (unemployment rate) + β_7 (air quality index) + δ_1(Liberal) + δ_2 (Conservative) + δ_3 (South European) + δ_4 (East Asian) + δ_5 (East European) + δ_6 (Others) + ε

Table 7-3. Life Expectancy, Welfare Regimes and Health Determinants

Independent Variables	Dependent Variable:	Life Expectancy			
	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	80.76*** [0.52]	79.52*** [1.60]	78.34*** [1.32]	81.67*** [1.89]	80.61*** [1.70]
Liberal Regime	– 0.82 [0.54]	1.01 [0.59]	1.21* [0.52]	0.90 [0.60]	1.09* [0.53]
Conservative Regime	0.13 [0.56]	1.07 [0.57]	1.57** [0.50]	1.13 [0.58]	1.66** [0.51]
Southern European Regime	1.18* [0.58]	2.84*** [0.71]	3.73*** [0.55]	2.25** [0.76]	3.28*** [0.59]
East Asian Regime	1.30* [0.56]	3.60*** [0.62]	3.94*** [0.54]	3.51*** [0.63]	3.84*** [0.55]
East European Regime	–4.00*** [0.65]	–1.35 [0.72]	– 1.05 [0.66]	– 1.69* [0.74]	– 1.35* [0.68]
Others	–5.36*** [0.63]	–1.35 [1.01]	– 1.68* [0.73]	– 2.12 [1.08]	– 2.38** [0.81]
log(GDP per capita)		5.67*** [1.04]	4.92*** [0.84]	5.21*** [1.08]	4.23*** [0.92]
Air Quality		–0.09*** [0.02]	– 0.10*** [0.02]	– 0.09*** [0.02]	– 0.11*** [0.02]
School enrolment ratio		–0.02 [0.01]		– 0.02* [0.01]	
Unemployment rate		0.01 [0.03]		0.02 [0.03]	
Disposable income Gini		–20.11*** [2.68]	– 17.6*** [2.5]		
Market income Gini				– 21.0*** [2.8]	– 18.3** [2.5]
tax and transfer effects				15.5*** [3.4]	13.0** [3.3]
Observations	282	244	256	235	247
R^2	0.6113	0.6995	0.7597	0.7067	0.7646
Adjusted R^2	0.6028	0.6852	0.751	0.6908	0.7546
F Statistic	72.08***	49.09***	86.43***	44.56***	76.66***

p<0.001***,<0.01**,<0.5*

We have six health indicators for Y in the three models. Table 7-3 shows the regression models with life expectancy as the dependent variable. As the first step, this researcher runs Model II in Table 7-3 to find some 'influential data points' which means "an observation that either by itself or along with other observations, has a demonstrably larger impact on the calculated values of various estimates (coefficients, standard errors, t-values, etc) than is the case for most of the other observations" (Belsley et al., 1980, p. 11; cited in Bollen and Jackman, 1985, p. 511).

The influential data point is a different concept from 'outliers' of which definition is "observation that are distinct from most of the data points in a sample" (Bollen and Jackman, 1985, p. 511). Outliers can have adverse effects on the multiple regression. Osborne and Overbay (2004) summarise that the outliers can 1) increase error variance and reduce the power of statistical test, 2) decrease normality altering the odds of making errors, and 3) bias the estimates. Not all outliers are necessarily influential data points unless their omission definitely leads to a significant change in the estimated parameters. In this context, there is no reason to omit outliers unless the data is wrongly coded.

However, if their figures distort individual parameter estimates and the overall fit of the statistics, then it is problematic. Therefore, this researcher detects the influential points, not outliers. The statistical program R has the 'influence.measures' function, which returns a rectangular array of five diagnostic functions to detect influential points (Kleiber and Zeileis, 2008, p. 99). The five functions are DEFIT (to measure the change of the fitted value after the deletion of the observation), COVRATIO (to measure the change in the estimate of the OLS covariance matrix after the deletion), hat values (the distance of the X values for the case from the means of the X-values for all n cases), Cook's distance (to measure how much the entire regression function changes: a measure that combine the standardised residuals and hat values in a single calculation.)

and DEBETA (to measur the change in the coefficient after the deletion) (Kleiber and Zeileis, 2008, pp. 95~100; Gordon, 2012, pp. 486~487).

The Influence.measures function highlights observations that are unusual for at least one of the functions. For example, for Model 2 in the third column in Table 7-3, a total of 32 regions turn out to be influential at least one of the five diagnostic functions. However, we don't have to lose all the information as there do not seem to be general rules on deciding what influential points should be deleted (Wooldridge, 2008, pp. 316~321; Hair et al., 2005, p. 246). The deletion may "result in removing too many participants to the point that the analysis can no longer be performed" (Cousineau and Chartier, 2010, p. 66). In this thesis, this researcher deletes exceptional cases whose influences are so extreme that their influences are statistically significant in more than two of the five diagnostic functions. In the case of Model 2, the three regions turn out to be the extreme cases. They are US California, France's Ill de France and Italy's Lombardy. Their selection is partly due to their large populations with respectively their 1st, 16th and 22nd largest number of their residents among the total of the subject 292 regions.

The multiple regression is run with the data weighted by the region's population. In the following multiple regression models, the influential points would be deleted in the same influence measure function and criteria. With the omission of the three extreme cases, 289 observations remain. In some circumstances, the number of cases should be reduced further. When a region's data for a variable is unavailable, it is treated as a missing value. For example, seven regions such as Canada's Yukon and Chile's Atacama don't have their life expectancy data. Therefore, the number of available cases is 282 for the variable (282 = 292 – 3 - 7). That is why we have the number of observations on the second column of Table 7-3. Likewise, the number of cases differs depending on the number of missing values in each model.

This researcher also examines possible violations of the assumptions for the multiple regression. In the case of Model 2, Shapiro-Wilk test rejects the null hypothesis that the model's residual is normally distributed ($W=0.9616$, p-value = 4.147e-05). The studentized Breusch-Pagan test also rejects the null hypothesis of homoscedasticity of the residuals ($BP=39.269$, p-value = 4.768e-05). The two tests mean that the model violates the assumptions of normality and homoscedasticity. The other models also commit the same violations. Given that this researcher cannot operationalise some theoretical health determinants such as behavioural factors and institutional factors (e.g. healthcare systems), it is obvious that they are 'omitted variables' (Wooldridge, 2008, pp. 84~85) probably hidden in the residuals. The omitted variables are regarded as affecting the normality and heteroscedasticity of residuals. These violations of the assumptions are inevitable due to the data unavailability.

In addition, the regression specification error test (RESET) (Ramsey, 1969) shows the evidence of functional form misspecification (RESET=3.50, p-value=0.03), hinting at probable missed important nonlinearities between the variables. It is another limitation of the models even though this researcher linearises one curvilinear relation by taking logarithm for the GDP per capita variable.

This researcher also tests the possible multicollinearity, the correlation among the independent variables. The high multicollinearity is problematic, because it can enlarge the population variances of the distribution of their coefficient and in turn raise the risk of obtaining erratic estimates of the coefficients (Dougherty, 2011, p. 165). A test of the variance influence factors (VIF) for the ratio variables among the independent variables are all lower than 4, far below the common cut-off threshold of a VIF value of 10 (Hair et al., 2005, p. 254). In Model 5, the VIF for the welfare effects and market income were 3.6 and 4.7 as well, avoiding the multicollinearity problem. The other following models with different dependent variables have the similar patterns of violations, assumed to be due to the omitted variables.

Table 7-3 demonstrates the outcome of the five multiple regression models with life expectancy as the single dependent variable. The following Table 7-4~8 also shows the almost same set of models, but the only difference is that each table has the different dependent variable. In the table, Model 1 on the second column has only the dummy variable to compare the weighted mean values of the five welfare regime in addition to the ‘others’. It is basically the same statistical outcome with the ones we see in Table 7-2 in that we compare welfare regimes in health outcomes. The ‘others’ here indicate the weighted values of health indicators of regions in Israel, Turkey, Mexico and Chile. They don’t form any distinct welfare regime group and consequently the ‘others’ indicator has only limited meanings just as a weighted average of this heterogeneous group.

In Table 7-3, Model 1 shows the regime differences in life expectancy. As this researcher sets the Scandinavian life expectancy as the ‘reference category’ (Dougherty, 2011, pp. 231~232), we cannot see its name on the first column. Instead, the intercept indicates the regime’s life expectancy estimate, 80.76 years. Without any control variables, the Scandinavian regime is estimated to have higher life expectancy than the Liberal regime and lower life expectancy than the Conservative regime, but the difference is not statistically significant. However, East Asians and South Europeans have longer (statistically significantly) lives than the Scandinavians. The estimated difference with the East Asian regime is 1.30 year. On the other hand, the East European regime is the only regime of which weighted average life expectancy is shorter than the Scandinavians in a statistically significant way.

In Model 2, this researcher adds the five control variables, which are selected as the aggregate-level health determinants in Chapter 5. In this case, the intercept does not have any particular meaning as this figure is estimated on the improbable condition of all the independent variables assumed to be zero. However, the coefficients of each welfare regime are the estimated difference in life expectancy in comparison with the Scandinavian indicator. Again, its differences with the East Asian and South European welfare regime are statistically significant. In addition, the estimated differences are almost tripled from 1.3 years to 3.6 years in the East Asian case.

Among the control variables, GDP per capita, air quality and disposable income Gini have a significant relationship with life expectancy. For example, a one percent increase of GDP per capita is associated with a 0.0567 year increase in life expectancy. Life expectancy is also expected to rise by 2.011 years in accordance with a 0.1 point drop in disposable income Gini index. Likewise, a one microgram increase in the air pollution indicator, pm2.5, is associated with a 0.09 year decrease in longevity. The influences of school enrolment rate and unemployment rate are counterintuitive in that the less educated and the more unemployed, the longer life expectancy is. However, their association is statistically insignificant.

Model 3 then omits the two insignificant control variables to see if it enhances the statistical fit of the model. With only the three statistically significant control variables remaining, the model shows clearer differences in life expectancy between the Scandinavian regime and the others. East Asians are estimated to live approximately 4 years longer than the Scandinavians, which is the biggest difference. All the other regimes have the statistically significantly longer life expectancy than the Scandinavian welfare regime except for the East European regime. There is no significant difference in life expectancy between East Europeans and the Scandinavians.

As seen in the previous chapter, Scandinavian nations have the lowest disposable income Gini, (0.2603) and the highest GDP per capita (44,255 dollar). All the five Scandinavian nations are also ranked within the top seven nations in terms of environmental quality in OECD Better Life Index (OECD, 2015a). Given these excellent health determinants of the region, it is not surprising that the Scandinavia's life expectancy would be lowered if the favourable variables are controlled for.

What is surprising is that even when this researcher doesn't control for the key variables, the Scandinavian model does not excel in life expectancy and, in fact, has the statistically significantly lower life expectancy than those of East Asian and South European welfare regimes as presented in Model 1. Therefore, controlling for all of their advantageous variables of high income, narrow income inequality and clean environment, their life expectancy is reduced below the level of those of the Liberal, Conservative, South European and East Asian regimes. The puzzling findings for the egalitarian welfare model would be further discussed in the next chapter.

Table 7-4. Female Life Expectancy, Welfare Regimes and Health Determinants

	Dependent Variables:	Female Life Expectancy			
Independent Variables	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	83.03*** [0.56]	82.42*** [1.57]	81.98*** [1.33]	83.37*** [1.86]	82.66*** [1.72]
Liberal Regime	− 1.06 [0.58]	0.69 [0.62]	0.94 [0.55]	0.65 [0.63]	0.91 [0.56]
Conservative Regime	0.59 [0.60]	1.28* [0.61]	2.10*** [0.53]	1.32* [0.63]	2.15*** [0.55]
Southern European Regime	1.64** [0.61]	2.53*** [0.74]	4.11*** [0.58]	2.26** [0.79]	3.99*** [0.62]
East Asian Regime	2.38*** [0.60]	4.64*** [0.65]	5.06*** [0.57]	4.60*** [0.66]	5.04*** [0.58]
East European Regime	− 2.48*** [0.68]	− 0.10 [0.75]	0.50 [0.68]	− 0.24 [0.77]	0.42 [0.71]
Others	− 5.13*** [0.66]	− 1.27 [1.01]	− 1.64* [0.75]	− 1.62 [1.08]	− 1.84* [0.83]
log(GDP per capita)		5.94*** [1.02]	4.02*** [0.84]	5.76*** [1.07]	3.82*** [0.92]
Air Quality		− 0.09*** [0.02]	− 0.12*** [0.02]	− 0.09*** [0.02]	− 0.12*** [0.02]
School enrolment ratio		− 0.03** [0.01]		− 0.03** [0.01]	
Unemployment rate		0.05* [0.03]		0.06* [0.03]	
Disposable income Gini		− 20.7*** [2.6]	− 16.7*** [2.5]		
Market income Gini				− 21.1*** [2.7]	− 16.9*** [2.5]
tax and transfer effects				18.5*** [3.4]	15.2*** [3.3]
Observations	279	241	253	232	244
R^2	0.665	0.753	0.7881	0.7546	0.7885
Adjusted R^2	0.6576	0.7412	0.7803	0.7411	0.7794
F Statistic	90.0***	63.47***	100.4***	56.11***	86.85***

p<0.001***,<0.01**,<0.5*

Model 4 tests whether 'market income Gini index' and 'tax and transfer effects', the conceptually related categories to disposable income Gini, may have independent influence on life expectancy. As presented in the table, they have statistically significant impacts on aggregate health. The 0.1 point increase in market income Gini, indicating the worsening income distribution, is expected to result in a 2.1 year decrease in life expectancy.

On the other hand, a 0.1 point rise in tax and transfer effects works in opposite way by increasing life expectancy by 1.55 year. We can estimate that re-distributional effects and market income inequality have independent effects on the aggregate health indicator. The school enrolment rate also has the statistically significant association with life expectancy but in a counterintuitive way. In Model 5, this researcher deducts education and unemployment variables from Model 4. Scandinavian life expectancy has a statistically significant difference with all the other welfare regimes, again higher only than that of East Europeans and below those of all the other models.

Table 7-4 illustrates the statistical outcomes on the relationships between the welfare regime, health determinants and female life expectancy. This researcher again detects the influential points after running a regression with Model 2 in the third column (Kleiber and Zeileis, 2008, p. 99). Three regions are omitted for their extreme values. They are France's Il de France, US California and Italy's Lombardy. The total number of observations is reduced to 289 cases.

Model 1 on the second column shows the differences in female life expectancy between the six groups. Again, the intercept on the first line is the estimated female life expectancy of the Scandinavian welfare regime, 80.03 years. Scandinavian females live longer than their counterparts in the East European regime by statistically significant differences. However, their life expectancy is statistically significantly shorter than East Asian and South European women. The differences are 2.38 years (with East Asia) and

1.64 years (with South European). There are no significant differences of the Scandinavian regime when compared with the Conservative and Liberal regimes.

When this researcher controls for the five health determinants (GDP per capita, air quality, school enrolment ratio, unemployment rate and disposable income Gini) in Model 2, the differences between the groups are wider. The estimated differences in female life expectancy between the Scandinavian and East Asian regimes are enlarged up to 4.6 years. Women in the Conservative regime begin to show the significant difference with Scandinavian females. The Scandinavian regime's difference with East European regime is not statistically significant any more. The intercept no longer has any meaning because it is based on the improbable assumption that all values of the variables are zero.

All the ratio variables are significantly related to female life expectancy. A one percent rise in GDP in capita is associated with an increase in female life expectancy of 0.0594 years. Female life expectancy increases by 2.08 years in accordance with a 0.1 decrease in disposable income Gini and by 0.09 year in step with a one microgram drop in pm2.5, the air pollution indicator. The outcomes on school enrolment rate and unemployment are statistically significant but they are counterintuitive because high unemployment and low education is associated with longer female life expectancy.

The close examination of the data explains some reasons behind these puzzling associations. This is partly due to the characteristic patterns of South European and Liberal welfare regimes. As seen in Figure 6-1, South European regions have the highest unemployment rates and the vividly lowest school enrolment rates but they have unexpectedly good health indicators. In contrast, the Liberal regime's regions have relatively low unemployment rate and the highest school enrolment rate, but their health outcomes are mostly the worst. Further studies are needed to analyse these puzzling associations between the two independent variables and health indicators.

In the simpler Model 3, the differences in female life expectancy between the welfare regimes are widened further. The estimated gap between Scandinavian and East Asian women's life expectancies is longer than five years when the three ratio variables are controlled for. The Scandinavian welfare regime fails to show any significantly better health than any other welfare regime. The three health determinant variables have the significant associations with the female health indicators. A one percent increase of GDP per capita is expected to lead a 0.0402 year increase in life expectancy. Life expectancy is also expected to rise by 1.67 years in step with a 0.1 point drop in disposable income Gini index. A one microgram drop in pm2.5 is related with a 0.12 year increase in longevity.

When this researcher splits the disposable income Gini index into the market income Gini and tax and transfer effects in Models 4 and 5 with or without the education and unemployment variables, both the market income distribution and the redistribution indexes have stable associations with the dependent variables. The East Asian welfare regime constantly excels in female health by the biggest difference with its Scandinavian counterpart. In Table 7-5, we can see male life expectancy as the single dependent variable. Scandinavian males show relatively better health compared with their females as seen in Table 7-4. Life expectancy here is statistically significantly higher than those of the Liberal and East European regimes without controlling for any other health determinants in Model 1, and is higher than even those of Conservative and East Asian regimes, but the difference is not statistically significant.

Table 7-5. Male Life Expectancy, Welfare Regime and Health Determinants

	Dependent Variables:	Male Life Expectancy			
Independent Variables	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	78.72*** [0.68]	77.13*** [1.94]	75.27*** [1.57]	80.18*** [2.27]	78.39*** [2.02]
Liberal Regime	− 1.50* [0.70]	0.73 [0.76]	1.09 [0.66]	0.64 [0.76]	0.97 [0.67]
Conservative Regime	− 0.50 [0.72]	0.34 [0.75]	1.08 [0.64]	0.51 [0.75]	1.29* [0.65]
Southern European Regime	0.53 [0.74]	2.14* [0.92]	3.54*** [0.70]	1.29 [0.98]	2.93*** [0.74]
East Asian Regime	− 0.38 [0.73]	2.17** [0.81]	2.74*** [0.70]	2.13** [0.81]	2.67*** [0.71]
East European Regime	− 5.85*** [0.82]	− 2.94** [0.91]	− 2.53 ** [0.82]	− 3.34*** [0.93]	− 2.86*** [0.84]
Others	− 5.75*** [0.79]	− 0.87 [1.25]	− 0.96 [0.91]	− 1.89 [1.31]	− 1.86 [0.99]
log(GDP per capita)		7.46*** [1.24]	6.47*** [1.01]	6.82*** [1.28]	5.51*** [1.10]
Air Quality		− 0.07** [0.02]	− 0.09*** [0.02]	− 0.08** [0.03]	− 0.10*** [0.02]
School enrolment ratio		− 0.03* [0.01]		− 0.03* [0.01]	
Unemployment rate		0.01 [0.03]		0.03 [0.03]	
Disposable income Gini		− 26.6*** [3.2]	− 23.5*** [2.9]		
Market income Gini				− 28.0*** [3.3]	24.5*** [3.0]
tax and transfer effects				20.1*** [4.1]	17.2*** [3.9]
Observations	276	238	250	229	241
R^2	0.518	0.6515	0.6983	0.664	0.7072
Adjusted R^2	0.5072	0.6346	0.687	0.6454	0.6945
F Statistic	48.18***	38.42***	61.71***	35.57***	55.56***

p<0.001***,<0.01**,<0.5*

When this researcher controls for the five health determinant variables, the inter-regime difference emerge again in statistically significant ways. East Asian and South European male life expectancies are significantly higher than the Scandinavian figures throughout Models 2~5 with the only exception for South European males in Model 4. In general, the health of Scandinavian males are on par with those of Liberal and Conservative regimes but statistically significantly better than those from East Europe.

When compared with female health outcomes in Table 7-4, the coefficients of the four income-related indicators (GDP per capita, two income Gini indexes and tax and transfer effects) are generally larger for male life expectancy than for female life expectancy, which hints that the male health is more sensitive to the direct materialistic condition than female health. In addition, the counter-intuitive relationship between unemployment and life expectancy disappears when it comes to male health. On the other hand, the life expectancy differences between regimes are apparent and wider for females than for males. For example, East Asian women live longer than Scandinavian females by 2.38~5.06 years but the differences are reduced to -0.38~2.74 years for male groups.

The association between air quality and health is stronger for females than males. We may infer that the female health can be affected relatively more by non-materialistic conditions such as environmental factors or some unobserved differences between welfare regimes, while males are more directly affected by the material factors.

The next indicator is the infant mortality rate. In Model 1, the intercept indicates the Scandinavian data of 2.69 deaths among their 0~1 year age group per 1,000 live births. Only East Asian babies die less frequently than the Scandinavians, but the difference is not statistically significant. The Scandinavian figure is significantly lower than those of the Liberal and East European regimes.

Table 7-6. Infant Mortality, Welfare Regimes and Health Determinants

	Dependent Variables:	Infant Mortality Rate			
Independent Variables	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	2.69*** [0.56]	– 5.33*** [1.15]	– 4.22*** [0.79]	– 4.74*** [1.38]	– 3.05*** [1.03]
Liberal Regime	4.00*** [0.57]	0.91 [0.62]	1.18* [0.56]	0.87 [0.63]	1.01 [0.57]
Conservative Regime	0.78 [0.60]	0.11 [0.61]	0.08 [0.51]	0.12 [0.62]	0.05 [0.51]
Southern European Regime	0.49 [0.61]	– 0.55 [0.74]	– 1.08 [0.56]	– 0.75 [0.80]	– 1.53* [0.61]
East Asian Regime	– 0.41 [0.61]	– 2.81*** [0.66]	– 2.43*** [0.56]	– 2.82*** [0.67]	– 2.59*** [0.57]
East European Regime	1.49* [0.69]	0.64 [0.75]	0.78 [0.61]	0.54 [0.78]	0.44 [0.64]
Others	5.38*** [0.85]	0.09 [1.02]	0.76 [0.91]	– 0.17 [1.07]	0.05 [0.99]
GDP per capita		– 0.023* [0.010]	– 0.004 [0.009]	– 0.025* [0.011]	– 0.012 [0.010]
Air Quality		0.027 [0.023]		0.023 [0.024]	
School enrolment ratio		0.013 [0.010]		0.013 [0.011]	
Unemployment rate		– 0.07** [0.03]		– 0.07** [0.03]	
Disposable income Gini		30.6*** [2.8]	27.4*** [2.7]		
Market income Gini				30.4*** [2.9]	27.3*** [2.8]
tax and transfer effects				– 32.5*** [3.6]	– 31.9*** [3.6]
Observations	268	247	252	238	243
R^2	0.6091	0.7578	0.7419	0.7598	0.7471
Adjusted R^2	0.6001	0.7465	0.7334	0.747	0.7373
F Statistic	67.77***	66.85***	87.31***	59.3***	76.48***

p<0.001***,<0.01**,<0.5*

When this researcher controls for the different sets of health determinant variables in the other Models 2~5, East Asia welfare regime's figures are outstanding. It is the only regime to show the significantly lower figures than the Scandinavians for all the models by the margin of more than two infant death cases.

Little significant differences could be found between all the other welfare regimes with only two exceptions of the Liberal regime in Model 3 and the South European regime in Model 5.

Of the health determinant variables, the three income inequality indicators are constant and significant determinants of the infant health, illustrating all the statistically significant impacts for all the models. Air quality was not statistically significantly associated with the dependent variable in Models 2 and 4. The environmental variable is consequently omitted in Models 3 and 5. The unemployment rate shows again the counterintuitive association with the infant mortality: the higher unemployment rate, the lower infant mortality rate. A partial explanation is that the Liberal regime with one of the lowest unemployment rates has by far the largest infant mortality rate. Probably some other variables may impact the regime's exceptionally high infant deaths, but the unaccounted-for relationships in this thesis might influence the outcomes.

When the three insignificant or untenable variables of education, unemployment and air quality are omitted, we can get Models 3 and 5. One of the most remarkable outcomes in the two models is that GDP per capita, one of the most constant health determinants, does not have a statistically significant relationship any longer with infant health. The disposable income Gini is the only variable with stable and significant relationships over the models and, even when split into the two different indicators, they still have significant relationships with the health indicators regardless of presence of other independent variables in Models 4~5.

Table 7-7. Child Mortality, Welfare Regimes and Health Determinants

	Dependent Variables:	Child Mortality Rate			
Independent Variables	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	0.28*** [0.04]	– 0.03 [0.09]	0.05 [0.08]	0.002 [0.102]	0.09 [0.10]
Liberal Regime	0.24*** [0.04]	0.03 [0.05]	0.05 [0.05]	0.024 [0.048]	0.05 [0.05]
Conservative Regime	0.06 [0.04]	– 0.03 [0.05]	– 0.03 [0.05]	– 0.025 [0.046]	– 0.03 [0.05]
Southern European Regime	0.04 [0.04]	– 0.15** [0.06]	– 0.17** [0.06]	– 0.16* [0.06]	– 0.18** [0.06]
East Asian Regime	0.02 [0.04]	– 0.18*** [0.05]	– 0.16** [0.05]	– 0.18*** [0.05]	– 0.17** [0.05]
East European Regime	0.16** [0.05]	0.06 [0.06]	0.06 [0.06]	0.06 [0.06]	0.06 [0.06]
Others	0.59*** [0.05]	0.11 [0.08]	0.25*** [0.07]	0.09 [0.08]	0.23** [0.08]
GDP per capita		– 0.002* [0.001]	– 0.0008 [0.0008]	– 0.002* [0.001]	– 0.001 [0.001]
Air Quality		0.004* [0.002]	0.004** [0.002]	0.003* [0.002]	0.004* [0.002]
School enrolment ratio		– 0.002* [0.001]	– 0.003*** [0.001]	– 0.002* [0.001]	– 0.003*** [0.001]
Unemployment rate		– 0.002 [0.002]		– 0.002 [0.002]	
Disposable income Gini		1.99*** [0.22]	1.71*** [0.21]		
Market income Gini				1.98*** [0.22]	1.71*** [0.21]
tax and transfer effects				– 2.08*** [0.27]	– 1.83*** [0.27]
Observations	277	238	245	229	236
R^2	0.6542	0.6881	0.7575	0.6925	0.7607
Adjusted R^2	0.6465	0.6729	0.7472	0.6754	0.749
F Statistic	85.14***	45.32***	73.1***	40.53***	64.74***

In Model 1 in Table 7-7, the intercept shows that 0.28 annual cases of deaths per 100,000 population of the 0~14 year-old group in the Scandinavian welfare regime. Given the rare frequency of the child deaths, the variance of the data is relatively small (0.05) and the coefficients are small in figures. Scandinavian children are significantly less prone to die when compared with children from the Liberal and East European regimes. However, there are no significant differences observed between the Scandinavian, Conservative, East Asian and South European regimes in Model 1 without controlling for any other health determinants.

In Models 2~5, East Asian and South European children die less frequently than their Scandinavian counterparts by statistically significant margins after this researcher controls for the health determinants. Air quality and school enrolment ratios have statistically significant influence on the child death. This is the first time the education variable begins to have the theoretically expected relationship with the health indicator: the more educated people are, the less likely die their children are. The disposable income Gini index also has the strong and constant impact either alone or after split into the two indicators. The unemployment rate does not show any significant relationship with the children's mortality records. It is noticeable that the GDP per capita no longer has the statistically significant relationship when the unemployment rate is omitted in Models 3 and 5. Overall, the influence of the income exerted on health of the infant or child groups seem to be inconsistent.

When the focus is moved to the old-age mortality as the dependent variable, the outcome shows an interesting contrast with the younger-age death statistics. Four regions are omitted for their extreme values including U.S. California, Southern Canto in Japan, Lombardy in Italy, and the Mexican Federal District. Again, in Model 1 in Table 7-8 shows that 4.67 deaths are annually reported for every 10,000 Scandinavian population aged over 65. This is significantly higher than those of the Liberal, Conservative, South European and East Asian regimes.

Table 7-8. Old-age Mortality, Welfare Regimes and Health Determinants

Dependent Variables:	Old-age Mortality				
Independent Variables	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	4.67*** [0.12]	4.55*** [0.27]	4.44*** [0.18]	4.35*** [0.33]	4.24*** [0.25]
Liberal Regime	– 0.26* [0.12]	– 0.56*** [0.15]	– 0.46*** [0.13]	– 0.55*** [0.15]	– 0.44*** [0.13]
Conservative Regime	– 0.44*** [0.13]	– 0.78*** [0.14]	– 0.75*** [0.13]	– 0.79*** [0.15]	– 0.76*** [0.13]
Southern European Regime	– 0.46*** [0.13]	– 0.97*** [0.18]	– 0.86*** [0.14]	– 0.90*** [0.19]	– 0.80*** [0.15]
East Asian Regime	– 1.21*** [0.13]	– 1.54*** [0.16]	– 1.42*** [0.14]	– 1.53*** [0.16]	– 1.41*** [0.14]
East European Regime	0.51*** [0.15]	– 0.09 [0.18]	– 0.09 [0.16]	– 0.04 [0.18]	– 0.05 [0.17]
Others	– 0.05 [0.14]	– 0.95*** [0.24]	– 0.58** [0.18]	– 0.86*** [0.25]	– 0.48* [0.20]
GDP per capita		– 0.01*** [0.002]	0.008*** [0.002]	0.010*** [0.003]	0.007** [0.002]
Air Quality		0.033*** [0.005]	0.032*** [0.005]	0.033*** [0.006]	0.032*** [0.005]
School enrolment ratio		– 0.003 [0.002]		– 0.003 [0.003]	
Unemployment rate		– 0.008 [0.006]		– 0.009 [0.006]	
Disposable income Gini		2.06** [0.66]	1.12 [0.59]		
Market income Gini				2.19*** [0.68]	1.23* [0.61]
tax and transfer effects				– 1.45 [0.84]	– 0.48 [0.79]
Observations	288	249	265	240	256
R^2	0.6007	0.6143	0.6031	0.6169	0.6057
Adjusted R^2	0.5922	0.5964	0.5891	0.5966	0.5896
F Statistic	70.46***	34.31***	43.06***	30.46***	37.63***

p<0.001***,<0.01**,<0.5*

The estimated margin is highest with East Asian at 1.21 death cases. Scandinavian data is lower only than that of the East European regime. However, once the health determinant variables are controlled for in Models 2~5, the statistically significant difference with the ex-communist regime also vanishes.

GDP per capita and air quality matter for the old-age health. The two variables have constant and statistically significant impacts on the older generation's health throughout Models 2~5. On the other hand, neither school enrolment ratio nor unemployment rate shows significant relationships with the elderly health indicator. Disposable income Gini index has the statistically significant relationship in Model 2 but it vanishes in Model 3. Models 4~5 may explain the reason of this inconsistent impact of the disposable income inequality indicator. For elderly people, their health is statistically significantly associated with market income Gini index, but not with tax and transfer effects. This is an unexpected outcome because elderly people, mostly after their retirement age, may benefit from the tax and transfer effects, such as pension, but few of them still earn incomes from markets. However, market income is also regarded as related more closely to individual status, as people can feel valued by the level of their market income (Rowlingson, 2011).

It is also noteworthy that the pension-related spending take the majority of the government welfare spending, for example in the United Kingdom (Hood and Johnson, 2014). The amount of each pension is in general proportional to individual's market income before their retirement because state pension works as "a transfer of money across individuals' lifetimes, rather than between different individuals in the cohort" (Crawford, Keynes, and Tetlow, 2014, p. 1). However, this account does not fully explain why old-age mortality is statistically significantly associated only with market income distribution, not with tax and transfer effects.

In Table 7-8, the addition of the ratio variables increases the value of R^2 only marginally, from Model 1 (0.6007) to the other models (0.6057~0.6169). It is mainly because of the higher number of observations included in Model 1 than those of the other models, but in part due to relatively less variation accounted for by the health determinants. Provided with the clear differences in old-age death rates between welfare regimes and the little contribution of additional variables in improving goodness of fit, we can infer that there might be unobserved variables accounting for the different outcomes between the regimes. In fact, when the regression model 2 is run without the dummy variables, its squared R is 0.084. It means that the four variables combined account for only around 8 percent of the variation of the old-age mortality.

This contrasts with the younger generation's health outcomes. For example, when this researcher applies the same model to the child mortality rates and infant mortality rates as dependent variables, the values of the squared R is as high as 0.4799 and 0.5294 respectively. In the case of the child mortality rate, the higher GDP per capita and the lower disposable income index are statistically associated with the lower children's death frequency. For infant mortality, the lower disposable income inequality and the higher school enrolment are related to the lower frequency of baby deaths. Overall, the health determinants, especially GDP per capita and income inequality indicators, have clear and significant influence on the younger generation's health, but not so much for the older generation.

7-4. Conclusion

In search for the answers to the research questions, the OECD regional dataset is used, for the first time in the cross-national comparative health studies. With the large set of the 292 regions, the multiple regression is conducted with the six dummy variables and seven independent variables. The following five findings emerge.

First, the Scandinavian welfare regime underperforms in all of the six health indicators. When controlling for some of the health determinants such as GDP per capita and income inequality, we observe their health indicators are statistically significantly worse than those of the South European and East Asian regimes. Especially, in terms of old-age mortality, its data is worse than those of the Liberal, Conservative, South European and East Asian welfare regimes. The statistical outcomes corroborate the presence of the second Scandinavian puzzle on its relatively low level of aggregate health.

Second, the East Asian welfare regime, generally ignored in the international comparative health studies, shows significantly good health outcomes. Its aggregate health is better than that of the Scandinavian welfare regime in statistically significant ways in the three health indicators: life expectancy, female life expectancy and old-age mortality rate. The differences between the two regimes are enlarged further when the health determinant variables are controlled for. In the other three health indicators (male life expectancy, child mortality rate, infant mortality rate), there are not statistically significant differences observed between the two regimes. However, when this researcher controls for the health determinants, East Asians again excel their Scandinavian counterparts again. This finding again supports the East Asian puzzle regarding its unexpected good aggregate health. We can infer that there are other strong health determinants that enhance East Asian health, which seem to be absent in the Scandinavian regions.

Third, among the health determinants, GDP per capita and the three income inequality indicators have strong and consistent association with aggregate health indicators. GDP per capita is associated with all the health data with only some exceptions for some models regarding the younger-age aggregate health. The three income inequality-related indicators also have strong and consistent relationships with all the health indicators with only two exceptions, 1) disposable income inequality and old-age mortality and 2) tax and transfer effects and old-age mortality. In general, income is relatively more strongly related to the older age mortality while income inequality is more to younger age health. We can observe these peculiar patterns involving the age threshold effect. Air quality is also a significant health determinant but not for infant mortality. Findings on school enrolment ratio and unemployment is challenging as they are associated with health indicators in counterintuitive ways. This finding needs further analysis in the future studies.

Fourth, we also observe different patterns between gender and age, corroborating the hypothesised thresholds in the aggregate health in the between-regime statistics. In the case of gender, female life expectancy is significantly higher in East Asian and South Europe than in Scandinavian regions, but the pattern is not significant any longer when it comes to male life expectancy. In addition, male life expectancy is relatively more sensitive to materialistic variables such as income and income inequality but female life expectancy is more associated with other factors such as air quality. The age threshold is also observed. For example, the Scandinavian regime shows the relatively good records in younger age mortality but its old-age mortality is one of the worsts. As discussed, the older generation's health seems to be more related to income, while infants' and children's health are relatively more susceptible to income inequality.

Fifth, with regards to 'GDP per capita threshold effects' discussed in Chapter 3, we could not find any convincing evidence to confirm the presence of the threshold. The income threshold is closely related to the Wilkinson Hypothesis (Wilkinson, 1992) positing that, in rich countries over a certain degree of income per capita, aggregate health is no more related to economic growth. In contrast, the findings in this chapter demonstrate the strong and consistent association between GDP per capita and aggregate health. It should be also noted that in the process of selecting the cases, this researcher screens out relatively poor OECD regions until the correlation between GDP per capita and life expectancy is not statistically significant any longer, to make a pool of the rich regions over a certain threshold. However, it does not mean that the association between GDP per capita and aggregate health vanishes. To be precise, the correlation between the income indicator and aggregate health vanishes over a certain threshold, but it is only because other health determinants are not controlled for. Therefore, based on the finding, we cannot support the Wilkinson Hypothesis as long as the income threshold is concerned. The following chapter draws together and discusses the empirical findings of Chapters 6 and 7.

CHAPTER 8 DISCUSSION

8-1. Introduction

In the review of reviews and the systematic review presented in Chapters 3 and 4, we identify the second Scandinavian puzzle on the regime's underperformance in enhancing aggregate health. The time-series cross-sectional regression analyses and the cross-sectional multiple regression in Chapters 6 and 7 also suggest that GDP per capita is the consistent health determinant and that income inequality is also a statistically significant determinant in parts of the population. The two rounds of regression analysis also show that the Scandinavian welfare regime, despite relatively high GDP per capita and narrow income inequalities, underperforms in showing better aggregate health outcomes, corroborating the presence of the second Scandinavian puzzle. In contrast, the East Asian welfare regime, in spite of its relatively low-income level and large income inequality, shows the best aggregate health indicators, which this thesis suggests as the East Asian puzzle.

Given these findings, this chapter has three aims. First and second, this researcher attempts to find possible answers to the respective two puzzles. Third, this researcher discusses the implications of the findings in this thesis in relation to the Wilkinson Hypothesis. This chapter consists of four parts. The first part discusses the possible reasons behind the second Scandinavian puzzle. The second seeks the answers to the

East Asian puzzle. The third part discusses the Wilkinson Hypothesis before we move on the fourth part of a conclusion.

8-2. Discussion 1: Causes of the Second Scandinavian Puzzle

8-2-1. Dual Scandinavian Puzzles?

As noted earlier, the first Scandinavian puzzle concerns the underperformance of Scandinavian regime in narrowing health inequalities, while the second puzzle focuses on its underperformance in aggregate health. Comparative health researchers have mainly focused on the first puzzle (e.g. Dahl et al., 2006; Eikemo and Huijts, 2009; Hurrelmann et al., 2011, Bambra, 2011; Mackenbach, 2012) but have arguably overlooked the second puzzle despite the growing evidence provided by empirical studies as seen in Chapter 3 and 4. For them, the Scandinavian excellence in aggregate health seems to be taken for granted as the Scandinavians are expected to be "doing well in overall health outcomes" (Bambra, 2013, p. 713) and to have "enviable health profiles" (Raphael, 2014, p. 10). Consequently, Bambra (2013) even goes on to suggest that the Scandinavia's underperformance in showing the narrowest health inequalities (i.e. the first puzzle) is paradoxical for the two reasons:

"First, that there is an implicit expectation (or normative 'belief ') within public health circles that better general health outcomes should be accompanied by smaller health inequalities; and second, that following public health theory, the social democratic welfare states with their more extensive, generous and egalitarian universal welfare should have smaller health inequalities." (p. 713)

As pointed out, this 'implicit expectation' is based on the logic that the first and second puzzles are like two sides of one coin. According to Gravelle (1998)'s account, when income inequality is enlarged, this will logically leads to larger health inequality and lower aggregate health simultaneously. To put it very simply, if a millionaire takes 100 pounds from a poor person, the health benefit for the rich person is minimal while the impact on the latter can be detrimental. It will logically lead to widened health inequality as well as lowered aggregate health. Consequently, the Scandinavian 'good' aggregate health and 'bad' health inequalities are simply an illogical combination for many comparative health researchers.

In that context, this thesis succeeds in resolving one of Bambra's (2013) two paradoxical reasons, because this thesis finds that there is no 'mismatch' between Scandinavian health inequalities and aggregate health outcomes. In other words, they consistently underperform in both of the categories. Scandinavian states have consistent records in both relatively low aggregate health and relatively wide health inequalities: the dual Scandinavian health puzzles. Then the next question may be why the Scandinavian welfare regime does not succeed in showing the best aggregate health outcomes. This researcher provides two clues to this puzzle in the next section.

8-2-2. Two Accounts on Second Scandinavian Health Puzzle

This thesis begins by casting doubt over the general assumption that the Scandinavian nations have "enviable health profiles" (Raphael, 2014, p. 10) and in the end provides empirical evidence to suggest the second Scandinavian puzzle. Thus we can assume that there must be some other factors that significantly worsen Scandinavian aggregate health, offsetting the favourable effects of the majority of variables such as income, income inequality and environment. This researcher at first critically discusses five potential 'other factors' hinted at in the previous studies. Then two novel solutions will be suggested for the puzzle with supporting empirical evidence.

It needs to be emphasised that few studies have discussed the second Scandinavian puzzle and therefore there is little previous analysis to draw upon. However, there are commentators who examine the 'first Scandinavian health puzzle' which is Scandinavia's underperformance in having the narrowest health inequalities. We can apply many of their analyses on the first puzzle to the second puzzle because, as abovementioned, the two puzzles are logically interrelated (e.g. Gravelle, 1998). Given this, we can consider five potential 'other factors' suggested by commentators on the first Scandinavian puzzle.

First, there could be detrimental behavioural factors, such as smoking, which worsen health among the Scandinavian populations. Preston, Glei, and Wilmoth (2010) find that in Finland, smoking-related causes account for 18% of its total female deaths in 1955, while the ratio did not exceed 1% in any other nations. Crimmins, Preston, and Cohen (2011) also claim that mortality rates for lung cancer and respiratory diseases are higher in Denmark, the United States and the Netherlands when compared with other rich nations and the causes are estimated mainly to be smoking. The prevalence of smokers in the adult population was very high in Denmark (57.5%) in 1970 (Tapia Granados, 2013).

However, findings on the past records for Denmark or Finland cannot be applied to the contemporary Scandinavian states. According to Preston, Glei, and Wilmoth (2010), the Scandinavian figures on smoking-related percentage of female deaths (4% for Finland, 5 % for Sweden, 6% for Norway) are generally lower than those of other nations such as United Kingdom (16%), Japan (12%) and United States (22%). Given that two other Scandinavian nations, Denmark (18%) and Iceland (21%), have relatively higher figures, smoking may be the main cause of death in the Scandinavian regions, but it is questionable that cigarette consumption kills a much higher ratio of Scandinavian people than people elsewhere. On the other behavioural factors, the Scandinavians consume relatively less alcohol (Tapia Granados, 2013) and have a relatively healthy diet (Adamsson et al., 2011).

A second vein of thinking is that Scandinavian health was better in the past compared with other nations but other nations such as Southern Europeans (Tapia Granados, 2010; Regidor et al, 2011) have now caught up. A possible explanation for this convergence might be that human beings already have reached their biological limit (Olshansky, Carnes, and Désesquelles, 2001) and the Scandinavian population has reached the ceiling earlier than others. However, the Scandinavian states have not only been caught up with but have been overtaken by other nations as seen in Chapters 5 and 6. This ‘overtaking’ is observed in other empirical studies (Tapia Granados, 2013; Auger, Le Serbon, and Rostila, 2015).

Third, some contend that Denmark, and probably Finland, should be excluded from the analysis of health of the Scandinavian or Social Democratic welfare states (Mackenbach and McKee, 2013; Raphael, 2014). The reason can be the "split slightly between the highest performing countries of Sweden, Iceland and Norway versus Finland and Denmark who performed slightly less well" (Popham, Dibben, and Bambra, 2013, p. 3). Mackenbach and McKee (2013, p. 395) state that "All the Nordic countries, except Denmark, and the Netherlands have a very good performance". Raphael (2014) simply excludes Denmark from its analysis on the Scandinavian aggregate health issue by stating that "Denmark is not included in this article as a health promotion leader as its health profile is rather poor for reasons not really understood" (p. 7) but still argues that "Nordic welfare states' accomplishments must be celebrated and used as a basis for maintaining the public policies shown to be successful in promoting the health of its citizens" (p. 7).

Regarding the omission of this Scandinavian outcast state, the two points need to be highlighted. First, without Denmark, the second most populous Scandinavian nations, it is questionable to combine the other three or four North European nations as a group of 'Scandinavian' nations. Second, even if we agree on the omission of Denmark, the other Scandinavian health records do not meet the general expectation either. To take a simple example, its archetype Sweden has its life expectancy at 81.9, ranked as the eight out of 34 OECD members, behind Switzerland (Liberal), Japan (East Asian), Italy and Spain (South European), Iceland (Scandinavian), France (Conservative) and Australia (Liberal) (OECD, 2013a). With Norway ranked at 10th, Finland (21th) was placed on the bottom half together with Denmark (25th). By omitting Denmark or possibly Finland, the Scandinavian average can certainly be lifted, but not up to the top level. Consequently, the omission of one or two Scandinavian nations cannot be a convincing way of answering the second Scandinavian puzzle.

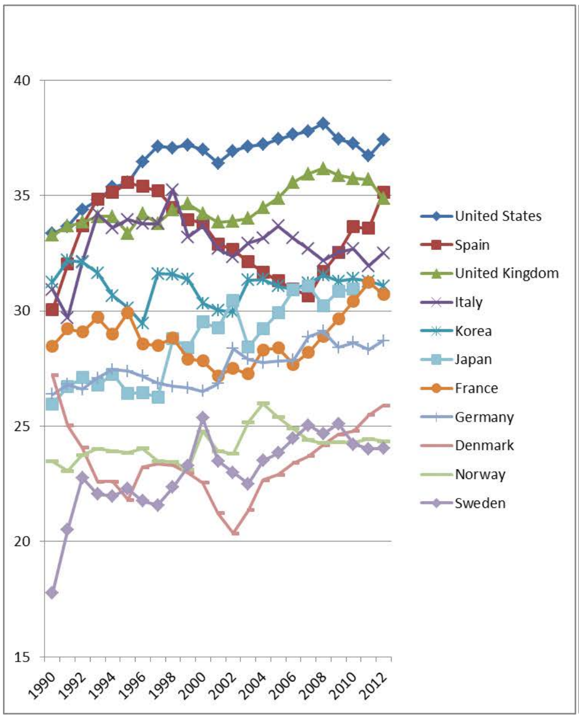
Fourth, Bambra (2011) suggests that the Scandinavian underperformance in narrowing health inequalities might be due to differences in "the data used and methods of measurement" (p. 742). This factor, which this thesis terms as artefact account III' in Chapter 2, can be applied to accounts for the second Scandinavian puzzle. Comparability of the dataset compiled by researchers or international organisations from individual nations has been a thorny issue. In the case of infant mortality, reporting regulations and practices differ between nations on counting the cases of infant death before registration (Sachs et al., 1995). The implication of this incomparability is significant and should not be overlooked, but we can hardly assume without evidence that the limitations lead to the unfavourable health outcomes particularly for the Scandinavian welfare regime.

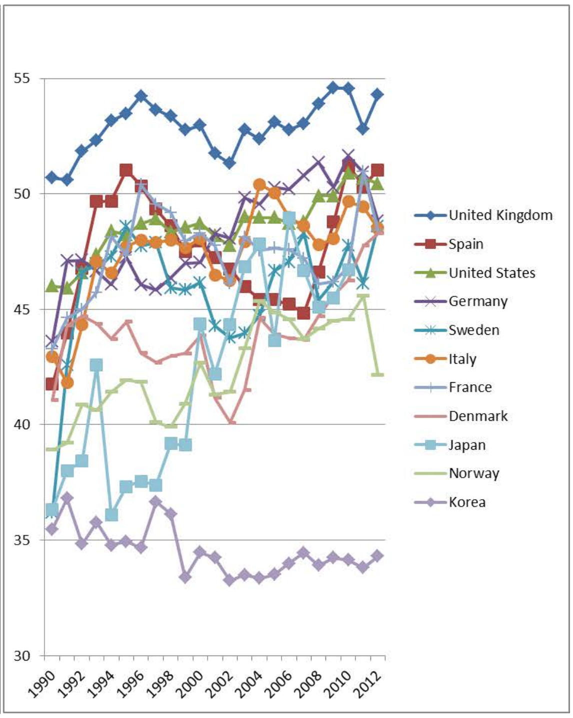
Fifth, the retreat, if not transformation, of the Scandinavian welfare states can influence its worsening aggregate health. The Scandinavian countries have seen their income inequalities simultaneously widen and poverty rates surge since the mid-1990s (Fritzell, Bäckman, and Ritakallio, 2012). Danforth (2014) goes on to argue that there is not a distinct Social Democratic welfare regime any longer, because the model, once salient between the late 1970s and the late 1990s, was merged with the Conservative model. Based on the chronological examination of the three welfare regimes with the two forms of cluster analysis over 18 welfare states from 1950 to 2000, he contends that the Social Democratic world, merged with Conservative one, has been succeeded by a 'European' world. The contention corresponds with the view that "the Nordic countries have become ‘more European’ with regard to income inequalities and poverty risk” (Kvist, Fritzell, Hvinden, and Kangas, 2012, p. 204).

However, Scandinavia’s move toward the European model can not necessarily mean that its welfare states deteriorate to a level in which it harms its aggregate health more than other welfare models. As Powell (2004) points out, “while there is general agreement that Social Democracy has changed in recent years, the parameters of change are less clear… the different possible changes include variation from the ‘ideal type’: variation from ‘old’ Social Democratic parties in office; convergence with other parties; and convergence between countries” (p. 2). For instance, Figure 6-6 shows the Scandinavian welfare regime’s variation from its old type and its convergence with other welfare types in terms of its income inequality. Its income distribution worsened markedly especially in 1995~2000, but still maintains income inequalities far below from those of the other welfare regimes. Despite its significant retreats therefore, Scandinavia’s distinct model has not been dismantled and maintains its over-performance in terms of its typical generosity and universality (Lindbom, 2001; Fritzell et al., 2012).

Given the review of some proposed explanation on the second Scandinavian puzzle, this chapter now suggests the two following original accounts, addressing relatively unexplored aspects of the Scandinavian model. First, the relatively high market income inequality of Scandinavian nations could be important. Figures 8-1 and 8-2 show the patterns of disposable income inequality and market income inequality for the 11 key industrialised nations, including three Scandinavian states plus two states respectively from the other four welfare regimes over the period 1990~2012.

Figure 8-1. Disposable Income Inequality Trend Figure 8-2. Market Income Inequality Trend





The nations on the legend of Figure 8-1 are placed in terms of their income inequality in the latest timing of 2012. The higher one's place is, the worse its income inequality is. At the top, United States, Spain and United Kingdom are located with their widest income inequalities. On the other bottom end, the three Scandinavian nations are clustered together. The locations of the three Scandinavian states, visibly far below of the other eight nations in the figure signifies the North European success in keeping the disposable income distribution to the lowest level, despite their contrasting trajectories over the period. Denmark draws a big 'U' curve meaning the fluctuation in income distribution with Sweden generally widening its income inequality. Norway maintains their income distribution fairly consistently.

Overall, the Scandinavian nations have kept a certain distance from all the other nations and their consistently low income inequalities lead health researchers to expect their aggregate health to excel in comparison with other welfare states (e.g. Bambra, 2011). The disposable (or net or post-tax or after-tax) income inequality indicators are commonly used by comparative health researchers because the statistics gauge the exact money pocketed by individual person or household after tax and welfare benefits (e.g. Wilkinson, 1992; Lynch and Kaplan, 1997; Mackenbach et al., 2008; Wilkinson and Pickett, 2009). This choice is justifiable because disposable income can "capture fully the effects of taxes, benefits" (Lynch and Kaplan, 1997, p. 305).

However, there is another aspect to be considered on the choice of the income inequality indicators. To examine this issue, we need to briefly go back to the controversy between the neo-materialists (e.g. Lynch et al., 2000) and the psycho-sociologists (e.g. Marmot and Wilkinson, 2001). The former group, focusing on the structural and physical determinants of health, criticise the latter group, stating that it is "hard to understand how this emphasis on psychological functioning and informal interpersonal relations would serve as a basis for a public policy agenda" (Lynch et al., 2000, p. 1204). On the other hand, the latter groups focuses on the "psychosocial effects of relative deprivation involving control over life, insecurity, anxiety, social isolation, socially hazardous environment, bullying" (Marmot and Wilkinson, 2001, p. 1234). This thesis is not concerned with adjudicating between these positions, but the differences in their stances have some implications in this context. Hypothetically, there can be a question: are two oranges essentially the same for an individual's health when the one is funded from a food stamp and the other from a salary after a week's delivery work? The answer from neo-materialists might be positive, but the psychosocial group's answer may be negative. Then the disposable income Gini might be arguably the best indicator for the neo-materialists but NOT for the psycho-sociologists because the post-tax indicator does not distinguish between earnings from market and those from welfare benefits. That is why Rowlingson (2011, p. 13) points out that "perhaps gross income is linked more closely to status, as people feel valued by the level of their gross income, while net/disposable income may be more closely linked to material differences in how much people have available to spend on material goods."

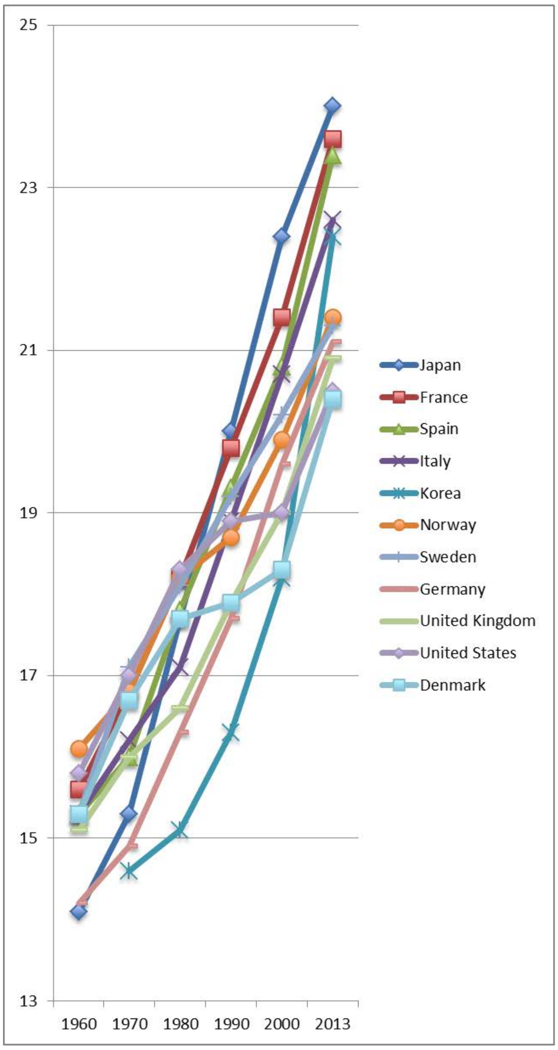
Given this possible validity of the market income inequality indicators, Figure 8-2 shows the cross-national trends over the last two decades. The legend box shows the different 'ranking' as of 2012. The bottom three nations are all changed except for Norway and the spots are replaced by the two East Asian nations. Japan's data are available only until 2010 and its 2010 indicator is compared with those of the other nations in 2012. The other two Scandinavian nations have their ranking moved up with Sweden even higher than Conservative France and South European Italy. In the case of the remaining two Scandinavian nations of Iceland and Finland, Iceland is placed as having the second fairest market income inequality just above South Korea, but the Finnish indicator is near to that of Japan. The readjusted places of the Scandinavian states in Figure 8-2 illustrate that the Scandinavian welfare regime succeeds in narrowing its disposable income gap with its generous welfare benefits or progressive taxation but not in suppressing the rising market income inequalities. Except for Norway and Iceland, the Scandinavian states are intermingled with other welfare regime states in the table. Consequently, the larger-than-expected disparity in market income can be another partial cause behind the second Scandinavian puzzle.

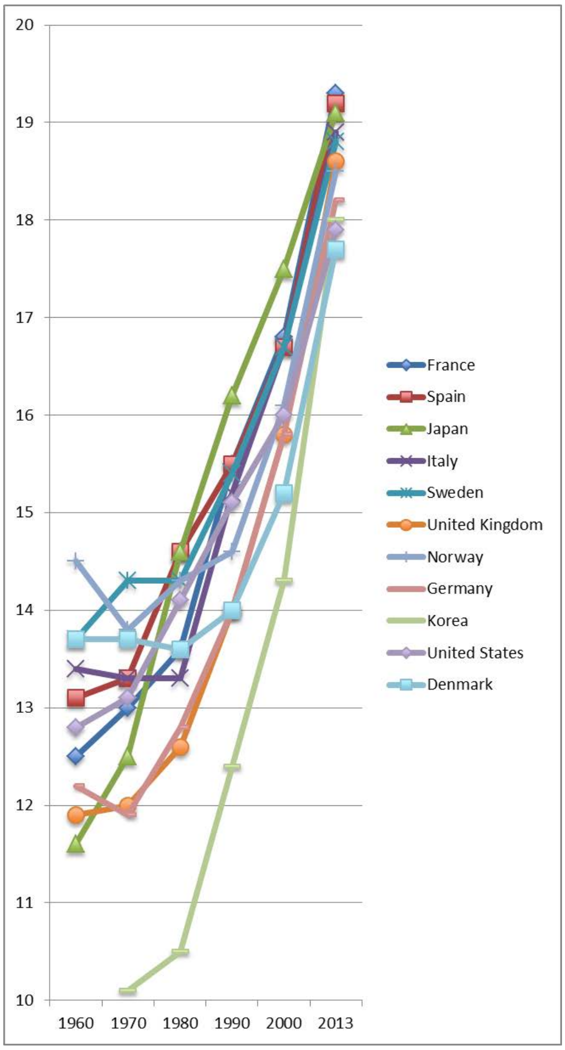
Second, old-age mortality rates may provide a clue to the second Scandinavian puzzle. We find in Chapter 7 that, of the various health indicators, the Scandinavia's disappointing old-age mortality rates are conspicuous, showing the worse outcomes than those of the Conservative, Liberal, East Asian and South European welfare regimes. The Scandinavian elderly indicators are better than the East European statistics, but even this difference was no longer statistically significant when other variables such as GDP or income inequality are taken into account. Figures 8-3 and 8-4 show the trends of female and male life expectancy at 65 of the 11 key nations, of which include two nations each from the four key welfare regimes plus the three Scandinavian nations of Sweden, Norway and Denmark.

The 11 nations on each legend box are listed in order of their measures from the best at the top to the worst at the bottom in the latest time of 2013. For example, in Figure 8-3, the top three nations of Japan, France and Spain show the longest female old-age life expectancies. At the bottom are located the three laggards: United Kingdom, United States and, notably, Denmark. The other two Scandinavian nations of Sweden and Norway are not located in the higher half of the list either. The location of the three Scandinavian nations again confirms their relatively poor records in old-age health indicators. What is more remarkable is their worsening ‘ranking’ over the decades.

In the 1960s, the Norwegian elderly were the longest living people with their remaining life expectancy at 16.1 years. Swedish and Danish elderly shared the fourth best spots together with Spain and Italy with life expectancy at 15.3. However, they have been overtaken by a number of nations over the decades. The simple calculation indicates that the three Scandinavian nations increased the elderly life expectancy by 5.5 years on average, which is only bigger that the average (5.3 years) of the two Liberal regime states (US and UK). The other regimes, with the average of the presented two nations, increased by 7.5 years (France and Germany), 7.7 years (Spain and Italy) and 8.9 years (South Korea and Japan).

Figure 8-3.Elderly Female Life Expectancy Figure 8-4. Elderly Male Life Expectancy





The elderly male health indicator, as presented in Table 8-4, illustrates an even sharper fall of Scandinavian health in terms of ranking. The three Scandinavian nations had the best male elderly life expectancy (14.5 years in Norway, 13.7 in Sweden and 13.7 in Denmark respectively), ahead of any other nations in the 1960s. The past five decades saw their statuses slide to the middle (Sweden as the fifth highest and Norway as the seventh highest), or even the bottom of the ranking (Denmark). The three nations only managed to increase elderly male life expectancy by 4.4 years on average, the worst accomplishment compared with the South European (5.8 years), Liberal (5.9), Conservative (6.4) and East Asian (7.7) welfare regimes.

These stagnant health profiles among Scandinavian elderly people have been observed in previous studies (Janssen, Mackenbach, Kunst, 2004; Staetsky, 2009; Rau, Soroko, Jasilionis, and Vaupel, 2008). Janssen et al. (2004) examine the old-age mortality in seven European nations in 1950~1999 and note that stagnation is observed in Denmark and Norway in the pace of decline in old-age mortality in the 1980s and 1990s, while France and England and Wales showed continuous, strong mortality decreases. They claim that smoking-related cancers and COPD (Chronic Obstructive Pulmonary Disease) contributed to the stagnation in Norway, but its impact in other nations is relatively modest.

Staetsky (2009), based on the analysis of female old-age mortality rates in nine Western nations plus Japan, contend that there are two contrasting patterns in changes of the health indicators between the mid-1980s and 2000 with the first group (France and Japan) showing a large decrease in mortality and the second (Denmark, the US and the Netherlands) lingering. The cross-national comparative analysis illustrates that "the divergence is, to a very significant extent, due to the differential impact of smoking related mortality" (p. 885). Focusing on the "countries with lowest mortality" (p. 764), Rau et al. (2008) observe the significant and universal decline in old-age mortality for the recent decades with some exceptions of the United States, the Netherlands and Norway.

The previous studies provide insight into the second Scandinavian puzzle but their focus is limited to Denmark or Norway and to the smoking factor as the single cause. The focus needs to be broadened, considering that "(e)ven after smoking related mortality is removed, differences in levels and trends of mortality persist, most remarkable are the differences between the Netherlands and Denmark with relatively high 'smoking free' mortality and slow pace of reduction during the last quarter of the 20th century, on one hand, and France and Japan, where 'smoking free' mortality was low and the pace of reduction was faster" (Staetsky, 2009, pp. 907~908).

With the empirical findings on the Scandinavian elderly health reviewed, we can draw three inferences regarding the Scandinavian old-age mortality. First, there are hidden or unidentified effects of a time-variant variable considering the Scandinavian old-age health, the best in the 1960s, has been dramatically relegated to one of the worst places. It is highly unlikely that all the other welfare regimes share some positive health determinants over the decades excluding only the Scandinavian states. If the hidden effects are time-variant in nature, it means that the effects are not the time-invariant factors such as weather, geographical atmosphere, or culture including diet. Second, the hidden effects may not be gender-specific because, as examined, the Scandinavian increases in elderly life expectancy are the worst for both of its male and female populations. Third, the Scandinavian welfare states, largely regarded as formulated since the 1960s (Danforth, 2014), have failed in counteracting the hidden effects, unless the welfare states themselves are a negative factor to their aggregate health.

To give clues to the second Scandinavian puzzle, this researcher proposes the two new potential factors. First, market income inequality in Scandinavian states is not as egalitarian as its disposable income inequality. The psychosocial effects of the larger market income inequality may influence people's feeling of accomplishment, financial independence and self-esteem, which may have negative impacts on its aggregate health.

Second, among generations, the significant stagnation of old-age health improvements in the Scandinavian states may account, at least partially, for the welfare regime's overall underperformance in enhancing aggregate health. However, causes of Scandinavian old-age mortality stagnation have been still largely unexplored.

8-3. Discussion 2: East Asian Health Puzzle

Chapters 6 and 7 show that the East Asian welfare regime has the best health indicators especially for female or old-age populations compared with the other welfare regimes. The East Asian health outcomes are so counterintuitive, given its relatively wide income inequality indicators and relatively low GDP per capita, that this thesis proposes it as the East Asian puzzle. In this section, this researcher briefly reviews some of the few comparative empirical studies that include the East Asian welfare regime, and discuss the possible causes behind the unexpected East Asian health outcomes.

Only three studies appear to include the East Asian welfare regime in their comparative cross-national health research (Karim, Eikemo, and Bambra, 2010; Chuang, Chuang, Chen, Shi, and Yang, 2012; Popham et al., 2013). Table 8-2 shows the summary of these three rare studies. We can see that the research designs are all different in terms of methods, case selections and methods. However, their findings indicate that the East Asian welfare regime show one of the best health outcomes compared with all the other welfare regimes by reporting the best life expectancy (Karim et al. 2010; Chuang et al. 2012) or the best later life mortality and best female life expectancy (Popham et al. 2013).

Table 8-1. Three Studies Including the East Asian Welfare Regime

	Karim et al. (2010)	Chuang et al. (2012)	Popham et al. (2013)
East Asian Nations included	Hong Kong, Japan, South Korea, Singapore, Taiwan	Hong Kong, Japan, South Korea, Singapore, Taiwan	Japan, Taiwan
Welfare regime compared	Six regimes	Six regimes	Seven regimes
Data	The World Factbook 2003	OECD, World Bank (80~06)	Human Mortality Database (2006)
Method	Descriptive, ANOVA	Multilevel model	Compares contribution of premature and later life mortality
How East Asians Fare	The East Asian regime had the highest average life expectancy, the third lowest IMR	In 2000~2006, East Asians have the best life expectancy record, and the second best IMR record	East Asians have the lowest later life expectancy
Conclusion (in abstract)	"the East Asian welfare states did not have the worst health outcomes" (p.45)	"East Asian welfare states did not have worse health than most welfare states" (p. e23)	No remark on East Asian health in the abstract

In particular, Karim et al. (2010) contains detailed information on the aggregate health of the examined 30 nations, including five East Asian territories of Japan, South Korea, Singapore, Taiwan and Hong Kong. This larger set of East Asian territories also shows the remarkable aggregate health records. Japan has the longest life expectancy at 80.9 followed by Singapore (80.4), Australia (80.1), Switzerland (80.0), Sweden (80.0), Hong Kong (79.9) and Canada (79.8). It is striking that the three territories, out of the top seven, are from East Asian. South Korea is a relative laggard in the continent with one of the lowest life expectancy at 75.36.

Japan's longevity has been for a long time discussed in cross-national health researchers (e.g. Marmot and Smith, 1989), but the findings here show that the trend can be applied to the overall East Asian region. It can be inferred that East Asians share some characteristics that enhance their aggregate health. This suggests that including other East Asian nations in this thesis would have reinforced the impressive East Asian health outcomes.

However, surprisingly, the three studies are reluctant or 'too modest' to accept the East Asian's health enhancement. Karim et al. (2010) only note in the abstract "the East Asian welfare states did not have the worst health outcomes" (p.45). Chuang et al. (2012) also conclude "East Asian welfare states did not have worse health than most welfare states" (e23). Chuang et al.'s (2012) conclusion is based on the finding that East Asian health outcome does not have statistically significant difference with those of the other welfare regimes in most of the cases after controlling for GDP per capita, age dependency and period effects. Popham, Dibben, and Bambra (2013) do not provide detailed comments on East Asian health outcomes as their research focus was on the Scandinavian health inequalities.

In the end, none of the studies fully accept East Asia's remarkable health outcomes. This thesis is arguably the first to identify and analyse East Asia's higher-than-expected health status by coining the counterintuitive finding as the 'East Asian puzzle' despite its "lower social, education and health expenditure as well as lower physician density" (Chuang et al. 2012, p. e25).

The next question is what can be the cause of this East Asian health puzzle? As no studies have previously identified the puzzle, there is no analysis on the issue. However, there are some, but not many, analyses on Japan's relatively long longevity (Marmot and Smith, 1989; Ikeda et al., 2011). Their analyses on Japan can give insights to the puzzle given the geographical proximity and cultural similarity between Japan and Korea. This thesis will critically examine their analyses and will provide its own original accounts for the East Asian puzzle.

First, the economic growth especially for the decades since the end of World War II may account for Japan's impressive health figures. For example, for the first decade since the war, Japan's life expectancy increased by 13.7 years (Sugiura, Ju, Yasuoka, and Jimba, 2010). In the case of South Korea, its rapid economic growth since the 1960s can account for the largest increase in life expectancy among OECD members (OECD, 2015f). Undoubtedly, the ascent to economic prosperity is one of the major factors behind the rapidly increasing aggregate health indicators of the two nations for the last decades, but it is still questionable if the economic growth can account for their outstanding health figures especially in comparison with Western states. As seen in Figure 6-2, the East Asian welfare regime's average GDP per capita (32.8 thousand dollars) is far lower than those of the Liberal (45.4 thousand), Scandinavian (44.3 thousand), and Conservative (38.9 thousand) counterparts. In other words, the East Asian economic growth might help Korea and Japan catch up with Western welfare states in terms of aggregate health, but could not explain the reason East Asians began to live longer than people in other welfare states.

Second, Japan's universal education system can contribute to its good health. Researchers point out "the early establishment of free compulsory primary education" (Ikeda et al., 2011, p. 1094) or "high enrollment rate in elementary schools and advancement rate to upper secondary schools" (Sugiura et al., 2010, p. 7) in accounting for Japan's longevity. The focus on the universal and quality education is not different in South Korea as "a higher level of educational attainment than other nations of the comparable per capita income" (Seth, 2002, p. 7). East Asia's "education fever" (Anderson and Kohler, 2013) may explain its rapid rise in life expectancy over the decades. However, one of few comparative cross-national data on education presented in Figure 6-2, shows that East Asia's tertiary school enrolment rate is the second lowest at 70 percent out of the five welfare regimes only above Conservative regime (59.51 percent). There may be more evidence needed to link education and longevity in East Asia.

Third, the East Asian diet may enhance people's health, especially in Japan (Marmot and Smith, 1989). The nation's healthy diet is assumed to play a positive part, contributing to the low rates of coronary heart disease and breast and colon cancer. This account might be applicable to its neighbouring Korea in consideration of the culinary similarity. For example, alcohol consumption is the lowest in the East Asian welfare regime compared with the other welfare regimes as illustrated in Figure 6-2 with Japan (7.3 litres) at the 5th lowest and South Korea (9 litres) at the 10th lowest among the compared 24 nations in 2010.

Fourth, the relatively narrow income inequality in East Asian nations is another favourable factor (Marmot and Smith, 1989). Wilkinson and Pickett (2009) frequently use Japan as a nation with the lowest disposable income inequality and the best health indicators in contrast to the United States. They use United Nations Development Programme Human Development Indicators. However, a controversy remains unsettled over Japan's inconsistent income distribution statistics (Ballas, Dorling, Nakaya, Tunstall, and Hanaoka, 2014). For example, OECD (2011) reports that Japan's income inequality is even wider than the OECD average. According to the Standardised World Income Inequality Database (SWIID), which is used in this thesis and regarded as the improved version of the other income inequality indicators especially for international comparability (Corneo, 2011; Bjørnskov, Dreher, Fischer, Schnellenbach, and Gehring, 2013), Japan shows the worse disposable income inequality (0.3095) than the average of the 24 nations selected in this thesis (0.2997) in 2010. Given these outcomes, the conception of Japan as the most egalitarian state needs to be called into question. South Korea's disposable income inequality indicators are not particularly egalitarian either, because its disposable income inequality is slightly worse than Japan (0.3139) in 2010 according to the SWIID.

Out of the four accounts of 1) economic growth, 2) education fever, 3) nutrition, and 4) narrow income inequality, the nutritional factor might arguably be the only undisputed contribution to the East Asian longevity. In addition, this thesis suggests the following three accounts for the puzzle.

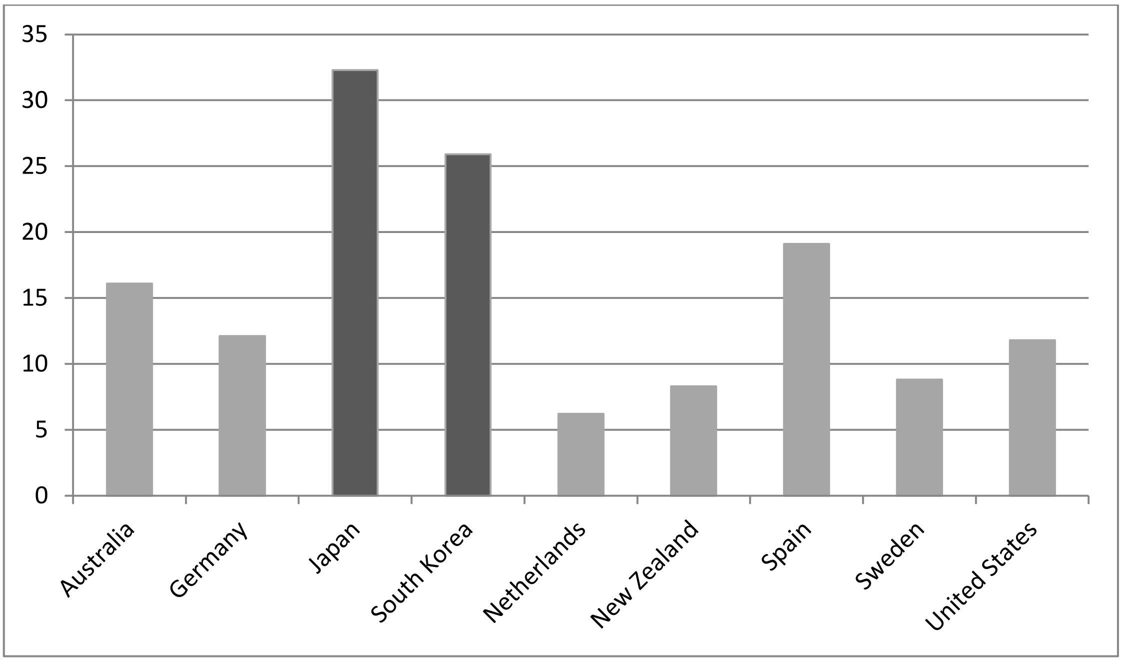
First, as discussed on the second Scandinavian puzzle, market, not disposable, income inequality seems to matter. Figure 6-2 illustrates that East Asian welfare regime's market income inequality is the lowest (0.4321) out of all the five welfare regimes. However, its weakest redistribution effects (0.1214) place the regime in the middle in terms of disposable income inequality indicator. Figure 8-2 also shows the two East Asian nations' places in terms of market income inequality with South Korea located far below the other nations in the figure and Japan also the third lowest just next to the second most egalitarian Norway. Market income can be more closely related to an individual perception of socioeconomic status than post-tax income (Rowlingson, 2011). The relatively fair market income distribution can exert significant psychosocial effects on East Asian aggregate health, even though the net income distribution is not as egalitarian as the pre-tax income distribution.

The second factor, which is relatively associated with East Asia's typical fair market income distribution, is the remarkably low unemployment rates in the two economies. The regime's average unemployment rate is only 4.68 percent, far below the second lowest Scandinavian data (7.33 percent) as seen in Figure 6-2. The 'worklessness', associated with poverty and social exclusion, has the detrimental effects on unemployed people (Harding et al., 2013; Bambra, 2011).

East Asian success in maintaining the low unemployment rates serves three aims simultaneously in the aggregate health policy perspective by saving state welfare spending for the unemployed (less need for redistribution), enhancing the market income distribution (boosting psychosocial health effects) and preventing the maleficent health effects on the workless group. The unemployment factor, together with the moderate market income distribution, may be one of the keys in solving the East Asian puzzle despite its state's low spending on welfare and public health.

Third, the elderly population of the East Asian welfare regime is especially healthier than those of any other welfare regime. In Chapter 7, 11 East Asian regions are located among the top 20 healthiest OECD regions for the aged. The average mortality rate of the regime is far lower than those of the other welfare regimes as shown in Table 7-8. The finding is consistent with a previous study comparing East Asia's outstanding old-age mortality with other regimes (Popham et al., 2013).

Figure 8-5. Ratio of Adults Living With Their Parents in Nine Nations



East Asia's emphasis on intergenerational ethics such as filial piety or respect for elderly might be a factor. Despite some studies on the East Asia's Confucius values (e.g. Sung, 2001), empirical evidence for East Asian value is "scarce" (Löckenhoff et al., 2015, p. 321) especially in comparison with those of other continents. Some researchers observe that the familial ethics and structure are growingly westernised, reducing the ratio of adult children supporting their elderly parents (Esping-Andersen, 1997).

However, East Asian intergenerational bonds remain relatively solid. According to the World Vales Survey Association (2015), responding to a question "do you live with your parents?", more than a quarter of respondents in Japan (32.3%) and South Korea (25.9%) said "yes", while the ratios of all the other seven nations were below 20%. The sixth wave (2010~2014) of World Values Survey covers 60 nations, out of which only nine nations coincide with the 24 nations analysed in this thesis. All the nine nations were presented in Figure 8-5. The relative high ratios in South Korea and Japan confirm the powerful familialism nested in the region, compared with other regions. Elderly people in East Asia are less likely to be vulnerable to loneliness or isolation, which is regarded as having a poisonous effect on aged population (Cattan, White, Bond, and Learmouth, 2005).

Consequently, East Asia's narrow market income inequality, low unemployment rate and the strong family bond may provide at least partial accounts for its unexpectedly good aggregate health. What needs to be added to the attempt to answer the East Asian puzzle is its incomprehensibly good female health indicator. In Chapter 6 and 7, we find that East Asian females live longer than those from any other welfare regimes. East Asian women's impressive health outcomes are also reported in other previous studies as well (e.g. Wilmoth, 1998). This finding is hardly reconcilable with the relatively low socioeconomic status of the East Asian women. In terms of Gender Gap Index published by the World Economic Forum (Hausmann, 2014), Japan and South Korea are ranked as the 104th and 117th, remote from the Scandinavian five nations clustered on the top five in the list. This counterintuitive finding poses another challenge for comparative health studies.

8-4. Discussion 3: the Wilkinson Hypothesis

In Chapters 6 and 7, we find that GDP per capita has statistically significant relationships with most of the aggregate health indicators across all the models. The disposable income inequality indicator has the expected association only with the infant mortality rate in the pooled time-series cross-section analysis in Chapter 6, and its association is statistically significant with all the health indicators in the multiple regression models in Chapter 7 with only one exception (Model 3 for old-age mortality).

Out of all the independent variables examined in the chapters, the two variables of income and income inequality have strong implication for the ongoing discussion on the 'Income Inequality Hypothesis' or the Wilkinson Hypothesis (e.g. Wilkinson, 1992, 1996; Avendano and Hessel, 2015). According to Pop, Ingen, and Oorschot (2013), the hypothesis states that "increasing societal wealth leads to improving population health only to a certain level of economic development. When this threshold of wealth is reached reducing disparities in income distribution is the key to further improve the health of the population" (p. 1028). As stated, the hypothesis consists of two sentences with the first on the contribution of wealth to health and the second on the contribution of income inequality. This researcher discusses the two sentences respectively because each has its own significant implication.

8-4-1. Income Contribution: GDP Per Capita Threshold

In the influential "*Spirit Level*", Wilkinson and Pickett (2009) note that as "living standards rise and countries get richer and richer, the relationship between economic growth and life expectancy weakens. Eventually it disappears entirely" (p. 6). In fact, this hypothesis corresponds with the 'GDP threshold effects' that this thesis observes and proposes in Chapter 4. Based on the hypothesis, this thesis selects the 26 national cases in Chapter 6 and the 292 regional cases in Chapter 7 by choosing the cases over a certain threshold of GDP per capita. This researcher sets a threshold, over which a statistically significant correlation between GDP per capita and life expectancy vanishes. For example, among the OECD members, the correlation is not statistically significant any more when the richest 29 nations over GDP per capita of 21,200 dollars in 2010 are selected.

On the other hand, we find that the GDP per capita is the most constant health determinants in both the pooled TSCS regression in Chapter 6 and the cross-sectional multiple regression in Chapter 7. It may sound paradoxical that income has statistically significant relationships with all the health indicators even though their correlation is not statistically significant within the selected set of cases. The reason for this is that the latter bilateral correlation test does not take into account unobserved variables, of which influences serve as 'noise' (Carmines, McIver, and others, 1981). The disappearance of the bilateral correlation among the affluent nations or regions may mean that other health determinants, possibly such as income inequality or education, exert relatively more influences on aggregate health among the rich cases than among the other less affluent cases.

The relatively reduced contribution of GDP per capita does not mean that economic growth stops working to enhance aggregate health. In other words, there is little evidence to support 'the GDP threshold effect'. Consequently, the solemn declaration such as "Economic growth, for so long the great engine of progress, has, in the rich countries, largely finished its work" (Wilkinson and Pickett, 2009, p. 5) might arguably be hasty given the continued and significant contribution of GDP per capita over time and across regions as found in this thesis. In other words, there is a certain threshold over which the income-health correlation is no more statistically significant, but it does not herald the end of health benefits of economic growth.

A closely related issue is the controversial limit of aggregate health improvement. Some researchers contend that life expectancy would not rise much further in the future (e.g. Olshansky, Carnes, and Désesquelles, 2001). Pop et al. (2013) also suggest the 'ceiling effect', stating that a further indefinite increase in life expectancy might not be physically possible. If the ceiling effect is at work, the additional wealth may also have limits in its contribution to longer longevity, which is compatible with the hypothesised end of the additional economic benefit.

However, as Oeppen and Vaupel (2002) contend, the limit is "broken" because the female life expectancy in the record-holding country has shown the steady rising trend for the last 160 years: almost 3 months per year. "In 1840, the record was held by Swedish women, who lived on average a little more than 45 years. Among nations today, the longest expectation of life—almost 85 years—is enjoyed by Japanese women" (p. 1029). In Chapter 6, we also find that Japanese, the longest living people, increased their life expectancy by 3.3 years in 1995~2010, higher than the average rise of either Liberal or Scandinavian welfare regimes (OECD, 2015f). Overall, the economic growth can be regarded as keeping enhancing aggregate health over and above the presumed ceiling.

8-4-2. Income Inequalities Determining Aggregate Health?, Gender and Age Thresholds

The second, and probably key, part of the Wilkinson Hypothesis is that "more equal societies were healthier because they were more cohesive and enjoyed better social relations" (Pickett and Wilkinson, 2015, p. 318). Based on the definition, a simple correlation test between income inequality and aggregate health may be enough to support the hypothesis for some researchers (e.g. Wilkinson and Pickett, 2009). However, other commentators contend that the relationship is spurious because the association vanishes after controlling for individual incomes (Gravelle, 1998; Jen, Jones, and Johnston, 2009), for GDP per capita (Mellor and Milyo, 2001; Engster and Stensoeta, 2011) or for educational attainment (Muller, 2002). For example, Muller (2002) conducts a multiple regression on aged adjusted mortality with independent variables of Gini coefficient and 'percentage of people aged over 17 years without a high school diploma' and finds that the income inequality effect vanishes when the educational variable is added to the regression model.

Given the complex and controversial pathways leading to aggregate health outcomes from the various health determinants, it is questionable to over-emphasise the role of income inequality in deciding aggregate health by positing that 'the more equal a society, the healthier it is'. This simple and tidy bilateral correlation may sound plausible and intuitively understandable, but cannot be free from criticism of its oversimplification of the complex web of health determinants leading to aggregate health. Income inequality can be an important predictor of aggregate health, but certainly not the only predictor. It follows that relatively equal societies in terms of income can have worse aggregate health than relatively unequal societies due to other health determinants.

That is the point where we can find keys to the second Scandinavian puzzle and the East Asian puzzle. The logics behind the expectation that the Scandinavian regime would have a relatively high level of aggregate health is that the welfare states' presumably egalitarian labour market and generous redistribution policies would boost aggregate health (e.g. Bambra, 2011; Hurrelmann, Rathmann, and Richter, 2010). When income inequality, especially disposable income distribution after taxation and welfare benefits, is not a single predictor of aggregate health, a relatively equal society is more likely to have relatively good aggregate health outcomes but this is not inevitably so. That is why we need to incorporate as many health determinants as possible in our analysis and avoid the simple bilateral correlation test in order to explain the complex web of health determinants.

Consequently, taking into consideration other health determinants, many, albeit not all, commentators go further to suggest that income inequality is not a statistically significant predictor of aggregate health. For example, Mackenbach (2002) in a BMJ editorial states that evidence favouring the income inequality hypothesis "has disappeared" (p. 1). Avendano and Hessel (2015) again report that "there is no strong evidence" (p. 597) for the hypothesis.

Despite this prevailing skepticism, the findings in this thesis do not support such death sentences on the income inequality hypothesis. The findings suggest a statistically significant association between income inequality and aggregate health of at least part of the population after adding all the available health determinants as control variables such as GDP per capita, school enrolment rate, unemployment rate and alcohol consumption (Chapter 6) or GDP per capita, school enrolment rate, unemployment rate and air quality (Chapter 7).

Even though this thesis identifies Scandinavia's relatively low level of aggregate health, it does not mean that we scrap the Wilkinson Hypothesis as well. To be precise, Scandinavian states do not succeed in having the best aggregate health 'despite' their favourable narrow disposable income inequality. In that respect, the findings here are closer to those of the proponents of the income inequality hypothesis or the Wilkinson Hypothesis (e.g. Subramanian and Kawachi, 2003; Pickett and Wilkinson, 2015) but further from those who expect the Scandinavian outperformance in aggregate health (e.g. Bambra, 2011; Mackenbach, 2012).

Some conditions need to be attached to this thesis's partial support of the Wilkinson Hypothesis. In Chapter 6, income inequality seems to wear a Janus face, because wide disposable income distribution is detrimental for infant health but, counter-intuitively, beneficial for old-age health of both genders. Unlike the inconsistent outcomes in Chapter 6, Chapter 7 finds disposable income inequality has negative relationships with all of the health indicators and the relationships are statistically significant except for one model on old-age mortality. From the findings in the two chapters, we can infer that old-age mortality is relatively unrelated or even positively related with income inequality, showing a remarkably different pattern with mortality rates of the other generations. Therefore, the findings in this thesis support the Wilkinson Hypothesis only for infant health, but not for the old-age health. In the end, this thesis has mixed findings for the Wilkinson Hypothesis. Without taking into consideration the age thresholds, we merely end up oversimplifying the impacts of income inequality on aggregate health.

8-5. Conclusion

In this chapter, this researcher discusses at length the two puzzles (the second Scandinavian puzzle and the East Asian puzzle) plus the Wilkinson Hypothesis in the light of the findings in Chapters 3~7.

First, on the second Scandinavian puzzle, this researcher examines the five possible accounts on the puzzle including 1) relatively high level of health hazardous behaviours such as smoking, 2) Scandinavia just having reached the ceiling in terms of aggregate health earlier and been caught up with by late-coming welfare states, 3) Scandinavian good health records without exceptional Denmark (or Finland), 4) artefact effects due to weak comparability of international data, and 5) retreat of Scandinavian welfare states. This researcher examines the meaning and limitations of the five individual accounts and suggests new two additional accounts for the puzzle. First, Scandinavian market income inequality is relatively wider than its disposable income inequality. Third, the relative stagnation of Scandinavian elderly mortality is especially remarkable over decades compared with the rapid decrease in the health indicators in the other welfare regimes. This stagnation awaits further analysis.

Second, on the East Asian puzzle, this researcher reviews the four possible accounts including 1) rapid economic growth, 2) compulsory education system, 3) healthy diet and 4) narrow disposable income inequality. At first, the first two accounts cannot be exclusively applied to the East Asian nations because they are common factors for developed nations. The third factor can be regarded as contributing to the East Asian health but, on the fourth factor, disagreement exists because some datasets categorise Japan and South Korea as relatively unequal states.

This thesis suggests the three additional and original accounts. First, East Asian market income inequality is the narrowest among the welfare regimes despite its weak redistribution efforts by the states. Second, its lowest unemployment rates among the five welfare regimes is also expected to prompt the enhancement in aggregate health by ‘catching three birds with one stone’: reducing state spending on welfare, narrowing the market income inequality and enhancing the health condition of otherwise unemployed people. Third, the still-strong familial bond in East Asian, possibly influenced by Confucius ethics and evidenced by related statistics, may bolster the elderly generation’s quality of life. With the highest ratio of intergenerational cohabitation in East Asia among the rich nations, aged people in the region may be less susceptible to loneliness or isolation, which may account for the lowest elderly mortality rates in Japan and South Korea.

Third, regarding the Wilkinson Hypothesis, this researcher divides the hypothesis into its two constituent parts, with the first on the role of income for aggregate health (economic growth no more contribute to aggregate health among richest nations) and the second on the role of income inequality (income inequality harms aggregate health among the richest nations). First, on the role of income over a certain threshold of income (i.e. the GDP threshold effects), we find that the GDP per capita is the strong and consistent contributor even among the richest group of nations or regions. This finding is the counterevidence against such contention as “the relationship between health and economic growth has levelled off” (Wilkinson and Pickett, 2009, p. 6) among rich nations.

Second, on the role of income inequality, despite the scepticism on the hypothesis (e.g. Mackenbach, 2002; Avendano and Hessel, 2015), the findings in this thesis partially support the harmful effects of income inequality with age threshold effects observed. The effects are statistically significant for infant mortality in both Chapters 6 and 7. On the other hand, the effects of income inequality on old-age health are relatively weak in Chapter 7 and even counterintuitive in Chapter 6. In the end, the thesis observes the age threshold effects on the relationship between income inequality and aggregate health.

In the end, we could still find the deteriorative effects of income inequality at least to part of a society's population. However, we can also find that income inequality, especially disposable income inequality, and GDP per capita play limited roles in accounting for the second Scandinavian puzzle and the East Asian puzzle. There seem to be more missing pieces in these two complex puzzles.

CHAPTER 9 CONCLUSION

9-1. Summary of Previous Chapters

After Chapter 1 set out the main themes and research questions of the thesis, Chapter 2 discusses the theoretical background of the Wilkinson Hypothesis and the second Scandinavian puzzle. In order to identify the main health determinants of aggregate health, it discusses eight accounts of health determinants: 1) artefact, 2) social selection, 3) materialistic and 4) behavioral/cultural, 5) psychosocial, 6) social capital, 7) environmental, and 8) policy. In addition, it reviews the discrete multi-level pathways along which each determinant influence individual, regional or national health. Finally it identifies the health determinants that both are theoretically robust and can be operationalised as independent variables.

Chapter 3, 'review of reviews' (Gough, Oliver, and Thomas, 2012) covers six previous review or systematic review articles which examine the relationship between income inequality, welfare regimes and aggregate health. This umbrella review method provides a wider perspective that shows the conflicting conclusions between (systematic) review articles despite their identical research questions. Their conclusions are inconsistent on both the two research questions of this thesis, showing a lack of consensus on the Wilkinson Hypothesis and Scandinavia's aggregate health. Finally, this chapter suggests three potential reasons behind the contrasting conclusions of the previous review articles.

Chapter 4 conducts a systematic review of the selected 48 primary articles. This chapter develops an original decomposition method, which breaks down individual articles into multiple findings based on their respective use of the four components: independent variable; dependent variable; methods; and; datasets. It shows that the reviewed empirical findings exhibit different patterns depending on age, gender, GDP per capita and period of the analysed population. This chapter suggests these particular patterns as 'threshold effects' over which the hypothetically expected relationships reverse or vanish.

Chapter 5 elaborates the methodological components that are subsequently used in Chapters 6 and 7: methods; case selections; independent, and; dependent variables. It should be noted that the case selection process in this chapter, arguably one of the most original elements of this thesis, is carefully designed and theoretically justified to select 26 nations out of the total 34 OECD membership and 292 regions out of the total 371 regions in the OECD regional datasets by using Pearson's product-moment correlation analysis (Chai, 2010, pp. 289~308). The method is used, based on the Wilkinson Hypothesis that the correlation between GDP per capita and aggregate health vanishes over a certain threshold of wealth of a state (Preston, 1975, Wilkinson, 1992)

Chapter 6 reports empirical findings based on the pooled TSCS dataset of 26 welfare states from five welfare regimes over 18 years. The five welfare regimes are Conservative, Liberal, Scandinavian, South European and East Asian ones as categorised in Chapter 5. First, the descriptive statistics of the nine independent variables and six dependent variables of the five welfare regimes illustrate that the Scandinavian welfare regime does not show the best aggregate health outcomes despite its favourable health determinants. In contrast, the East Asian welfare regime shows in general the best aggregate health outcomes among the welfare regimes compared despite its relatively unfavourable health determinants. Second, the TSCS regression with AR(1) Prais-Winsten correction and panel-corrected standard errors (PCSE) for the 26 nations for the 1995~2012 period demonstrates that GDP per capita is the most constant and significant health determinant. The other health determinants such as income inequality and alcohol consumption have limited influence on the health of only some population groups such as elderly people, infants or either gender group. Finally, the second Scandinavian puzzle and East Asian puzzle can also be observed in these chronological perspectives.

Chapter 7 reports empirical findings for the larger set of 292 regions within the OECD member states. A multiple regression analysis is conducted with the seven independent variables and six welfare regime dummy variables. The Scandinavian welfare regime underperforms in all of the six health indicators. When controlling for some of the health determinants such as GDP income per capita and income inequality, all of their health indicators are statistically significantly worse than those of the East Asian welfare regime. These statistical outcomes again corroborate the presence of the second Scandinavian puzzle. On the other hand, the East Asian welfare regime consistently shows one of the best levels of aggregate health among the welfare regimes. GDP per capita is again the constant and statistically significant health determinant of aggregate health of the independent variables. Unlike the previous chapter, income inequality indicators begin to show constant and statistically significant influences over health of all the population groups with only few exceptions. In addition, 'age threshold' or 'gender threshold' effects, suggested in Chapter 4, are observed. However, GDP per capita threshold effect cannot be observed as it has statistically significant associations with aggregate health among the rich regions.

Chapter 8 discusses the three key issues of the thesis: the 'Second Scandinavian puzzle', the 'East Asian puzzle' and the Wilkinson Hypothesis. First, this chapter suggests two possible accounts, which might explain the causes of the second Scandinavian puzzle. The first is that Scandinavian market income inequalities are not markedly low in comparison with those of the other welfare regimes. The third is that Scandinavian mortality of elderly people, once the lowest level among the welfare states, has stagnated and reached the highest level among the welfare states. Similarly, there are three possible accounts which might explain the causes of the East Asian puzzle. First, East Asian market income inequality is noticeably low. Second, East Asia has longstanding low unemployment rates. Third, its remarkably low elderly mortality indicators may be attributed to its highest ratio of intergenerational cohabitation among industrialised nations.

Finally, the chapter returns to the Wilkinson Hypothesis. This thesis divides the hypothesis into the two statements: one on economic growth and the other on income inequality. It refutes the first part of the hypothesis, because GDP per capita is a statistically significant determinant of most aggregate health indicators among the richest nations and richest regions. In other words, there is no 'GDP threshold effect' observed, over which the relationship between income and health is expected to reverse or vanish. On the second part of the hypothesis, the thesis has mixed findings, because the influence of income inequality is statistically significant in Chapter 7, but not in Chapter 6 except for infant mortality.

Overall, in response to the two research questions of the relationship between income inequality, welfare regimes and aggregate health, this thesis finds that aggregate health indicators of the Scandinavian welfare regime are worse than hypothetically expected while those of the East Asian regime are better than hypothetically expected. Given the level of GDP per capita, income inequality, educational attainment and environmental factors such as air quality, the health outcomes in the Scandinavian and East Asian regions are unexpected and counterintuitive, which this thesis proposes as 'the second Scandinavian puzzle' and 'the East Asian puzzle'.

The two puzzles, however, do not refute the relationship between aggregate health and the health determinants of GDP per capita, income inequality, educational attainment and environment. On the contrary, GDP per capita is the most consistent and significant health determinant in the majority of time-series cross-section models and multiple regression models in this thesis. Educational attainment is also one of the most powerful health determinants in the TSCS models. Income inequality has limited or sometimes counterintuitive statistical relationships with aggregate health in the TSCS models but its statistical associations are significant in the multiple regression, whether the indicators are market income or disposable income. This suggests that income inequality may not influence national-level aggregate health over the period 1995~2010, but its influence can be observed at the inter-regional level.

The impacts of the health determinants often differ along the age groups and between genders, corroborating the presence of age- and gender- thresholds. In addition, GDP per capita is a consistent and powerful health determinant among the selected rich 26 nations and 292 regions. There is a threshold over which the correlation between GDP per capita and aggregate health indicator is no longer significant, but there is *not* a threshold over which GDP per capita ceases to function as a statistically significant health determinant. This means that economic growth can be interpreted as contributing to further enhancement of aggregate health even among the developed nations, and this finding refutes the claim that economic growth ceases to enhance aggregate health in rich societies (e.g. Wilkinson, 1992, Wilkinson and Pickett, 2009).

Given the findings on the key variables, the current theories on the health determinants provide few clues to the second Scandinavian puzzle or to the East Asian puzzle. In other words, given the mostly favourable profile of its income, income inequality, educational attainment and environmental quality, Scandinavian aggregate health should be better than those of any other regions. The counterintuitive underperformance in terms of aggregate health, found in this thesis, reveals a critical limitation of current studies on aggregate health determinants.

This puzzle suggests that other unobserved health determinants are strong enough to counteract and overwhelm all the favourable influences of the tested health determinants. For example, in Table 7-4 in Chapter 7, the remarkable five-year difference in female life expectancy between the Scandinavian welfare regime (82.66 years) and the East Asian regime (87.70 years), when controlling for the other independent variables, can be interpreted as stemming from effects of unobserved variables. Without controlling for the independent variables such as income and income inequality, the gap is reduced to 2.38 years. In other words, Scandinavia's favourable variables such as high income and narrow income inequalities could narrow the gap in the health indicators from 5.04 years to 2.38 years.

However, the present theories have limitations in explaining what health determinants make the initially observed five-year gap between East Asian and Scandinavian females. In other words, within the frame of the current studies on aggregate health determinants, the theoretical expectation that Scandinavian health would be better than those of any other wealthy regions is justifiable, but it has certain limitations in explaining the large gap in aggregate health levels observed in this thesis.

East Asian aggregate health is the other extreme case, of which unobserved health determinants are strong enough to counteract and overwhelm effects of many unfavourable health determinants such as relatively low GDP per capita and weak welfare states. Similarly, within the current theoretical framework, the prediction of East Asia's poor aggregate health can be justified, but the framework could not simply explain the East Asian puzzle. Provided with the counterintuitive findings on both Scandinavian and East Asian health records, this thesis proposes two and three potential accounts for the two puzzles respectively as discussed in Chapter 8. They may provide some clues to the two puzzles but not enough to account for the unexpected large gaps in aggregate health levels between the two welfare regimes.

9-2. Original Contributions to Knowledge

9-2-1. Methodological Contributions

This thesis makes four methodological contributions. First, in Chapter 3, it uses a 'review of reviews' arguably for the first time, in cross-national comparative health studies. In a rare book by social scientists on the method of systematic review (Petticrew and Roberts, 2008), an umbrella review of prior systematic reviews has not been mentioned. As the authors note, "the science of systematic reviewing for social policy purposes is still relatively young" (p. xiv) and the method of 'review of reviews' remains an unfamiliar concept. However, review of reviews is an effective method to resolve the divergent, contradictory findings of prior systematic reviews. By taking a comparative analytical approach, we are able to identify key differences in inclusion, categorisation and interpretation practices that lead the reviewers to such differing conclusions. In addition, the new method also illustrates the limitations of previous systematic reviews such as oversimplifying evidence of primary articles even when the empirical studies provide multiple findings.

Second, this thesis introduces a decomposition review method in the systematic review in Chapter 4. To overcome the limitations of previous systematic reviews, as discussed in the review of reviews in Chapter 3, the new decomposition method deconstructs each article according to four components: independent variable; dependent variable; methods; and datasets. The selected 48 articles are broken down into the 107 findings after a coding process of the four components. This multidimensional approach enables us to identify an age threshold over which health outcomes show different patterns depending on the age of subject populations. For example, income inequality or welfare regime has the expected influence on health of younger population but not of older generations. In addition, three other thresholds effects involving gender, income and period can be observed.

Third, this thesis introduces a robust and theory-informed method for case selection in Chapter 5. In analysing the relationship between aggregate health and either income inequality or welfare regimes, the case selection criteria have not been clear in previous studies. Some set different level of GDP per capita as cut-off points such as 5000 dollars (Wilkinson, 1996) or 12,500 dollars (Lindstrom and Lindstrom, 2006). Others use different guidelines such as World Bank criterion (Ellison, 2002), OECD member nations (Macinko et al., 2004). The number of 'developed' nations fluctuates from 11 (Lindstrom and Lindstrom, 2006) to 40 (Vaupel, Zhang, and Raalte, 2011). The various sets of welfare states are found to be one of the reasons for different conclusions on the relationship (De Vogli, Mistry, Gnesotto, and Cornia, 2005; Pop, Ingen, and Oorschot, 2013). This thesis arguably avoids this case selection bias by conducting a correlation test between GDP per capita and health indicators such as life expectancy. This statistical test enables us to identify the GDP per capita threshold over which the correlation is no longer significant. This researcher selects the national or regional cases over the threshold, based on this systematic and theory-backed method.

Fourth, this thesis uses a regional dataset covering the majority of regions in OECD's 34 member states. The use of the regional dataset may resolve the chronic 'small-N' problem in the cross-sectional comparative health study. The nation-level datasets also have limitations in that they average out any regional characteristics. As Hudson (2012) points out, "when confronted by regional variations in policy, the major cross-national data sets offer a range of methodological 'fixes' to the problem: typically aggregating or averaging national data or selecting national exemplars or omitting non-standard regions" (p. 456). Many researchers have tried to overcome these limitations by focusing on regional differences in health but their analyses are confined to single nations, such as United States (Kennedy, Kawachi, Prothrow-Stith, and others, 1996; Lynch et al., 1998) or Japan (Shibuya, Hashimoto, and Yano, 2002). Regional-level studies which cross international borders are rare due mainly to lack of comparable datasets. Ross et al. (2005) is one exception as it compiles datasets from five different nations. OECD regional datasets (OECD, 2008; OECD, 2014d) provide a breakthrough as the dataset contains key health indicators such as life expectancy, infant mortality rate and crude death rate in hundreds of regions in the OECD member states, enough to be utilised for quantitative analysis. This thesis is arguably the first case to use these datasets in an international comparative health study.

9-2-2. Empirical Contributions

This thesis also makes five empirical contributions. First, its empirical findings across the chapters consistently demonstrate Scandinavia's counterintuitive underperformance in aggregate health. These findings contradict the general expectation by the majority of comparative health researchers (e.g. Wilkinson and Pickett, 2009; Mackenbach, 2011; Bambra, 2011). In Chapter 3, the review of reviews identifies the inconsistent conclusions of the three related systematic reviews. The systematic review in Chapter 4 also finds that a significant number of primary articles provide evidence on the presence of the second Scandinavian puzzle. Then the following empirical chapters also demonstrate that the Scandinavian welfare regime does not show the best aggregate health indicators. In contrast, it shows one of the worst health outcomes in terms of health of older generations in Chapters 6 and 7 among the welfare regimes compared. In Chapter 8, this thesis also suggests two possible reasons behind the second Scandinavian puzzle. The first is the relatively wide income inequality in the Scandinavian states especially in comparison with the traditionally narrow disposable income. The second is the stagnating old-age mortality reduction in the Scandinavian region over recent decades. The two accounts, together with other 'unobserved variables', seem to overshadow the favourable health effects of the Scandinavian regime's narrow income inequality, high income and clean environment.

Second, this thesis also appears to suggest significantly good health outcomes in East Asia, after identifying Japan and South Korea as forming a distinct welfare regime. This thesis is arguably the first case to acknowledge the regime's positive health outcomes and suggest it as 'the East Asian puzzle'. The findings in Chapters 6 and 7 also corroborate the presence of the puzzle. Finally, this thesis discusses three possible accounts for the puzzling findings. The first and second are the lowest levels of market income inequality and unemployment rate in the East Asian states. The third is the high ratio of intergenerational cohabitation of elderly parents and adult children, which may account for the lowest mortality rates in East Asian welfare regime in comparison with other welfare regimes.

Third, this thesis discovers that the relationships between income inequality, welfare regimes and aggregate health often reverse over a certain age threshold. The systematic review in Chapter 4 captures the trend that the majority of studies on infant and child mortality support the hypothesised relationship, but the empirical findings on working-and old-age mortality tend to repudiate the hypothesis. The pooled TSCS regression in Chapter 6 also shows that the negative relationship between income inequality and aggregate health is statistically significant for infant mortality rate when the income inequality indicator is the disposable income Gini. However, the relationship is reversed for the older population. This counterintuitive finding is also observed in some previous studies (e.g. Lynch et al., 2001; Muntaner et al., 2002). In Chapter 7, there are negative relationships between the disposable income Gini coefficient and all of the health indicators, with the only exception being one model on old-age mortality presented as Model 3 on Table 7-8. Overall, the systematic review and the empirical findings in this thesis suggests that the hypotheses involving income inequality, welfare regimes and aggregate health can be applied to younger-age populations but need more empirical evidence before they can be applied to older generations.

Fourth, by taking into account the gender- and age-threshold effects, this thesis has reached mixed empirical findings on the Wilkinson Hypothesis. The findings in this thesis support the hypothesis in the case of infant mortality rate, but not for the old-age health. This researcher remains cautious in respect of other health indicators such as life expectancy because the cross-regional regression supports the hypothesis but the pooled TSCS regression does not show statistically significant relationships. In that context, the conclusion in this thesis disagrees with the skeptical stance on the hypothesis such as the editorial of BMJ asserting that the evidence favoring the hypothesis "has disappeared" (Mackenbach, 2002, p. 1) or the commentary that "there is no strong evidence" (Avendano and Hessel, 2015, p. 597) for the hypothesis. This thesis does not wholly agree with the proponents of the hypothesis (e.g. Wilkinson and Pickett, 2009) for two reasons. First, its findings support the hypothesis only for infant mortality rate in Chapter 6 and for all the health indicators in Chapter 7. Second, the findings, unlike the hypothesis, support the role of GDP per capita for health even among the wealthy nations.

Fifth, this thesis affirms the long overlooked health determinant of market income inequality after testing its impacts on aggregate health. The majority of studies on the Wilkinson Hypothesis use the disposable income inequality indicator rather than the market income inequality because the former indicators measures the after-tax income actually pocketed by every household. Few studies use the market income inequality Gini with few exceptions (e.g. Sanmartin, 2003). However, as Rowlingson (2011) emphasises, “gross income is linked more closely to status, as people feel valued by the level of their gross income, while net/disposable income may be more closely linked to material differences” (p. 13). The pooled TSCS regression finds little evidence that market income Gini has a statistically significant relationship with any health indicators. However, the multiple regression in Chapter 7 finds that market income inequality has as statistically significant associations with aggregate health as the disposable income inequality does. The pre-tax income distribution also provides a clue for both the second Scandinavian puzzle and the East Asian puzzle as discussed in Chapter 7.

9-2-3. Conceptual Contributions

This thesis also presents three conceptual contributions, which are closely related to the empirical findings of this thesis. First, the new concept of ‘the second Scandinavian puzzle’ is arguably for the first time proposed in this thesis after a series of empirical findings revealing the lower-than-expected aggregate health outcomes in the North European welfare states. As discussed, the counterintuitively wide health inequalities in the Scandinavian welfare regime has been actively acknowledged and given several terms such as "Scandinavian paradox" (Richter et al., 2012, p. 860; Hurrelmann, Rathmann and Richter, 2010, p. 6) or "puzzle" (Bambra, 2011, p. 740), "paradoxical finding" (Huijts and Eikemo, 2009, p. 452). However, the new puzzling findings (e.g. Karim et al., 2010; Regidor, 2011) regarding Scandinavia's underperformance in aggregate health has long been overlooked, probably due to the prevalent expectation for, or belief in, Scandinavia's good aggregate health. The thesis also suggests another term of 'the dual Scandinavian puzzles' for Scandinavia's double underperformance in both aggregate health and health inequality. It should be also noted that the proposition of the new terms is not aimed at underestimating or disregarding beneficent effects of the egalitarian welfare states in the Social Democratic tradition. In fact, it is found in this thesis that Scandinavia's narrow income inequality, high educational attainment and good environment contribute to enhancing its aggregate health. The aim is to reveal the ungrounded 'beliefs' in Scandinavia's good aggregate health and to shed light on the reasons behind the puzzle.

Second, this thesis introduces the term of 'the East Asian puzzle' based on the empirical findings of the counterintuitively good health records in the East Asian welfare regime. Despite its impressive health indicators, the East Asian regime has been long overlooked in the scholarly literature with a few exceptions (e.g. Abdul Karim, Eikemo, and Bambra, 2010; Chuang, Chuang, Chen, Shi, and Yang, 2012; Popham, Dibben, and Bambra, 2013). Even these three studies seem rather reluctant to accept the unexpected outcomes. This reluctance might be due to the fact that East Asian nations have few positive health determinants such as their weak welfare states or wide disposable income inequalities.

Third, this thesis also suggests the term of ‘threshold effects’ involving age, gender, income and period. As discussed in Chapters 3 and 4, without taking into account this age-threshold effect, previous systematic review articles reach inconsistent conclusions by oversimplifying the dynamic pattern of the relationship between health determinants and generational- or gender- group aggregate health. We could observe in Chapters 6 and 7 that the each generational or gender health indicator often show different outputs in response to an identical input of health determinants. The findings in the chapters demonstrate the age-specific patterns that the hypothesised relationships between income inequality, welfare regimes and aggregate health are generally supported for younger populations but not for the adult or elderly generation groups. Consequently, a caution also needs to be exercised that none of infant mortality rate, working-age mortality, old-age mortality or either gender health indicator can serve as a representative indicator of aggregate health of a whole population.

9-3. Limitations

This thesis is not free from limitations. First of all, the omission of tobacco consumption as an independent or control variable is one of the critical limitations in the research design. As “a huge cause of ill health and premature death” (Wilkinson and Marmot, 2003, p. 24), smoking is estimated to be related to 2.7 million deaths from lung cancer, more than 7 million deaths from cardiovascular disease, and more than 14 million deaths from all forms of tobaccosis (Ravenholt, 1990). This strong determinant of health, if omitted, would serve as an unobserved variable and distort the statistical outcomes (Carmines, McIver, and others, 1981). The OECD datasets have tobacco data such as ‘annual consumption of tobacco items (e.g. cigarettes, cigars) in grams per capita aged over 15’ but it has too many missing values. For example, three nations of Austria, Belgium and Portugal have no statistics available in 1995~2012 and Canada, Spain and Sweden have some missing values along the period. Missing data under 10 percent of the observation can be generally ignored (Hair, Black, Babin, Anderson, and Tatham, 2005, p. 79), but in this case, the ratio of the missing data far exceeds the guideline.

The second limitation is the omission of another independent variable, namely wealth. In rich countries, wealth inequalities are regarded as wider than income inequalities. According to Nowatzki (2012), studies on income inequality report that Gini coefficients range around 0.2~0.4 in developed societies, but Gini coefficients for wealth inequality increase to around 0.5~0.9. In addition, the seminal work by Thomas Piketty (2014) warns that it is not labour income but gradually inherited wealth that divides the rich and poor in the 21st century. The total amount of private wealth is around two or three years of national income (200~300 percent) in developed nations back in 1970, but the capital/income ratio jumps to 400~600 percent in 2010 in rich nations such as the UK and France. In slowly growing economies, the relatively small flow of income makes little differences compared with past stock of wealth.

This “radically new structure of inequality” (Piketty, 2014, p. 32) may account for the inconsistent conclusions on the relationship between income inequality and aggregate health in previous literature, because the main factor dividing the haves and have-nots may have been shifted from income to wealth. Notably, “there is growing evidence that wealth might have an independent effect on people’s health” (Rowlingson and MacKay, 2012, p. xii). However, the majority of empirical studies, including this thesis, focus on the conventional labour income distribution.

Nowatzki (2012) is one of few studies examining the relationship between wealth inequality and aggregate health. After conducting bivariate cross-sectional analyses of the relationship between wealth Gini coefficient and aggregate health indicators such as life expectancy and infant mortality and controlling for aggregate-level confounders such GDP per capita, the author concludes that wealth inequality may be a stronger predictor of aggregate health than income inequality. However, the cases in the study include only 14 rich nations due to data unavailability. It is likely that wealth would become an important variable in the international comparative health research when a robust dataset becomes available.

Third, this thesis cannot be free from aggregation bias or the ecological fallacy. As explained in Chapter 3 and 4, this researcher could not employ individual-level health data because this thesis does not use self-rated health (SRH) survey outcomes. Consequently, regional and national-level aggregate health data are utilised in this thesis. This aggregate approach inevitably averages out individual differences in health, or ignores health inequalities within the population, especially health conditions of the disadvantaged.

However, it needs to be emphasised that this limitation is unrelated to the claim that the negative relationship between income inequality and health may be an ecological fallacy, when taking into account the curvilinear relationship between income inequality and health at the individual level (e.g. Gravelle, 1998; Jen, Jones, and Johnston, 2009). The thesis is not concerned with whether or not the negative income inequality-health relationship is only ‘spurious’ (Jen, Jones, and Johnston, 2009, p. 643) or an ‘artefact’ (Gravelle, 1998, p. 382). Whatever their claims may be on causes of the negative relationship, these authors unanimously agree that the relationship between income inequality and aggregate health should be negative. However, the empirical findings contradict this expectation. That is the point where this thesis begins its analysis. Therefore, this thesis may have a certain limitation as a ecological study, but is not related to the common claims on the ecological fallacy on the Wilkinson Hypothesis .

Fourth, this thesis proposes the four thresholds in analysing the relationship between income inequality and health: age, gender, GDP per capita and period. However, due to the word limit, the period effects could not be tested and only briefly discussed despite their potentially significant implications. The analysis of the pooled TSCS dataset in the third part of Chapter 6 seeks to show the chronological aspect of the relationship between income inequality, welfare regimes and aggregate health, but it is not sufficient to illustrate the dynamics over time.

Previous studies suggest that the relationships temporarily either emerge (Mellor and Milyo, 2001) or vanish (Wilkinson and Pickett, 2006). For example, Mellor and Milyo (2001) split their dataset by decade and find that Gini coefficient has a significant detrimental effect on health only in the 1970s and not in the 1960s, 1980s and 1990s. It follows that “the earlier cross-sectional findings reported by Rodgers, Flegg, and Waldmann cannot be reproduced when data from later time periods are used” (Mello and Milyo, 2001, p. 503). Wilkinson and Pickett (2006), after reviewing empirical articles, note that the relationship temporarily disappeared during the 1980s and early

1990s when, the authors note, income gaps widened particularly rapidly in many nations. They suggest three possible reasons for the temporary disappearance of the relationship. One of the three accounts is that the rapid decline in old-age mortality for the last decades may counteract the maleficent effects of income inequality with the timing of the trend earlier in some nations than others. Whatever the reasons, if this period threshold is observed, it needs further analysis.

In addition, there might be other thresholds as well. For example, another unexplored threshold is related to Gini coefficients. Kondo et al. (2012), after conducting a multivariate meta-regression analysis using 23 empirical studies on income inequality and health indicators, found the presence of a "threshold effect" (p. e11) of Gini over 0.3. "(I)nequality on health may exist even in less unequal societies but the impacts become stronger when income inequality exceeds a certain threshold value" (p. e14). The detrimental impact of income inequality particularly over the threshold is indeed significant, as it could provide a target for policy (Rowlingson, 2011). However, this thesis could not address this potentially significant point.

The fifth limitation is that the thesis could not provide convincing explanations for some counterintuitive evidence found in Chapter 6 and 7. There are two counterintuitive relationships that turn out to be statistically significant. The first is between the three income inequality-related indicators and old-age life expectancy in Chapter 6. The second is between some health determinants such as school enrolment rate, unemployment rate and life expectancy indicators in Chapter 7. From the first paradoxical finding, it can be inferred that the more unequal income distribution is, the healthier its elderly population is. This researcher may set up a hypothesis that people aged over 65 who outlive those who passed away earlier may be rich beneficiaries from the unequal wealth distribution, which make them healthier. However, given some findings that wide income inequality not only hurts the poor but also the rich (e.g. Subramanian and Kawachi, 2006), the hypothesis needs more empirical evidence. For the paradoxical findings in Chapter 7 on school enrolment ratio and unemployment, this thesis could not provide fully convincing accounts.

The sixth limitation involves use of some datasets. First of all, the SWIID (Solt, 2014), used to gauge the national-level income inequality in Chapter 6, imputes scores for missing values, providing convenient and complete datasets for all the 26 nations in focus. However, its imputation model to reduce missing values can incur potential bias when used in a statistical model (Jenkins, 2015), even though there is an ongoing controversy over the degree and impact of the imputation in the SWIID (See Jenkins, 2015, Solt, 2015). The quality of various cross-national income inequality datasets have recently been one of main subjects among the economic researchers (Atkinson & Brandolini, 2009; Ferreira, Lustig & Teles, 2015). The discrepancies between the international datasets are so large and obvious that, in one extreme case of a nation of Armenia, its redistributive effect is negative in SWIID (i.e., disposable income inequality is higher than market income inequality), while positive in another dataset called CEQ (Ferreira, Lustig & Teles, 2015). In other words, the limitation of this thesis of using the SWIID is directly related to the limitations of the current level of quality of contemporary cross-national economic dataset we have. In addition, the SWIID contains only Gini indicators and not other income inequality measures such as 90:10 ratios, thereby narrowing the focus of this thesis.

Another database of the OECD regional dataset, used in Chapter 7, provides a breakthrough in the international comparative health research by overcoming the chronic ‘small N’ problem. However the relatively new and pioneering database also reveals some limitations such as some apparently erroneous coding (e.g. crude youth death rates as seen in Chapter 5) or excessive data figures for some very small regions like the surprisingly high infant mortality rate in Canadian northernmost area of Nunavut. The region has the highest infant mortality rate of 16.2 deaths per 1000 live births. A single additional infant death case would increase the region's mortality rate dramatically with only some 30,000 residents in the region. Another issue is the small sample problem in gauging some measures such as poverty or income inequality in small regions. For example, in a small region such as Ceuta in Spain, the number of sample is as small as 113 household (Piacentini, 2014). Then "the reliability of estimates can be challenged for several small regions" (p. 11). It should be noted that the limitations of the datasets might lead to a bias in some of the statistical outcomes in this thesis. In the end, the datasets, used in this thesis, have obvious limitations, which in turn suggest that the conclusions are subject to some biases and should be revisited and revised in the light of new data as and when this becomes available.

Finally, the biggest limitation of this thesis may be that it raises more research questions than it resolves. In other words, this thesis may succeed in finding gaps in the previous studies by providing evidence to suggest the second Scandinavian puzzle and the East Asian puzzle. The findings might simply be a statement of what many researchers already know but are hesitant to publically acknowledge, as they run counter to conventional thinking or theoretical expectations. Given this, this thesis has been able to take only a few steps further to find answers to these puzzles. In Chapter 8, this thesis tries to draw some explanations but they remain mostly speculative and based on a few related statistics. In that respect, the limitations and merit of this thesis need to be critically reviewed.

9-4. Implications for future study

The abovementioned limitations of this thesis represent possible areas for future studies to focus on. There are also additional implications of this thesis for future research.

On the controversial choice of welfare regime as an independent variable accounting for national health differences, two systematic reviewers on the relationships between welfare regimes and aggregate health draw similar conclusions that the "black box" (Brennenstuhl et al., 2011, p. 399) welfare regime approach should be scrapped for "a multitude of different types of studies" (Bergqvist et al., 2013). Researchers should always seek better research designs to answer research questions but in this case we need not scrap the research question itself. In other words, the welfare regime approach may be scrapped for other possibly better approaches in international comparative health studies, but the question still remains why the Scandinavian nations do not show the best health outcomes in comparison with those of other populations, and why the East Asian nations show unexpectedly good health outcomes. The two puzzles strongly hint that there are unanalysed health determinants that overshadow the effects of the conventionally used health determinants such as income, income inequality and welfare states. (Even though this thesis attempts to identify some of unexplored health determinants in Chapter 8, they need more empirical evidences to be theorised.)

In that context, by seeking answers to the second Scandinavian puzzle and East Asian puzzle, this thesis at least partly breaks the 'alibi' of the unobserved variables. In other words, circumstantial evidence strongly hints at the presence of both 'unidentified killers' of Scandinavian populations and 'hidden invigorators' for East Asians. We cannot yet identify what these are, but, from the empirical evidence we found, we can infer that their influence on aggregate health in Scandinavia and East Asia are powerful enough to offset all the expected effects of health determinants such as income and income inequality. In these circumstances, the role of researchers is not to reject the welfare regime approach itself, but to respond to the theoretical challenges by seeking to identify the hitherto hidden causes of the puzzles. This process may help us to find 'missing pieces' of health determinants and to resolve the puzzles.

However, this thesis does not insist that the missing pieces can be found only within the welfare regime approach. For example, some findings in Chapter 7 also call for research on geographical or cultural characteristics of some regions, because some neighboring regions straddling the cross-national or cross-welfare regime borderlines show impressive health outcomes. Out of the top 20 regions with the highest female life expectancy, six are northern Spanish regions (Navarra, Madrid, La Rioja, Castile & Leon, Cantabria and Basque Country) and five are southern French regions (Pays de la Loire, Corsica, Aquitaine, Midi-Pyrénées and Rhône-Alpes). The common characteristics in the neighbouring regions across the borders might be one of the unexplored strong health determinants. These outcomes imply that either presumably natural or cultural characteristics exert as heavy an influence on health of the residents as political and policy factors do.

Table 9-1. Another East Asian Puzzle?

	Scandinavian welfare regime	East Asian welfare regime
Health Inequalities	I. The first Scandinavian puzzle: Health inequalities in the Scandinavian welfare regime are wider than expected.	III. ? (Another East Asian puzzle?)
Aggregate Health	II. The second Scandinavian puzzle: Aggregate health in the Scandinavian welfare regime is worse than expected.	IV. East Asian puzzle: Aggregate health in the East Asian welfare regime is much better than expected.

Given that this thesis suggests the East Asian puzzle and the second Scandinavian puzzle in addition to the first Scandinavian puzzle, some may recognise another gap requiring further investigation: probably another East Asian puzzle on health inequalities. Table 9-1 illustrates what this researcher has explored (cell I, II, IV) and what we have not explored (cell III).

While cell I (the first Scandinavian puzzle) is much discussed by comparative health researchers (e.g. Bambra, 2011; Mackenbach, 2012), this thesis focuses on the expanded research areas of cell II (the second Scandinavian puzzle) and cell IV (the East Asian puzzle). With the three puzzles (I, II, IV) examined, the remaining cell III on the East

Asian health inequalities has yet to be explored. Presumably, there might be two possible hypotheses: the first being that the East Asian welfare regime shows relatively wide health inequalities as theoretically expected, and the second that it shows counterintuitively narrow health inequalities. The first hypothesis (i.e. expectedly wide health inequalities in East Asia) may fit theoretical expectations but may not fit with the other three puzzles. In other words, cell III is the only non-puzzle surrounded by the three puzzles. On the other hand, the second hypothesis (i.e. unexpectedly narrow health inequalities in East Asia) may complete this tabulation of the four puzzles, potentially named as ‘the second East Asian puzzle’. This unexplored area also awaits empirical analysis.

9-5. Policy Implications

The findings in this thesis have the following implications for health policy. First of all, to relieve market income inequality may be a more efficient way to enhance aggregate health than to narrow disposable income inequality. We find that market income inequality, relatively ignored in overall emphasis of the other disposable income inequality, may have the equivalent or even more influence on aggregate health. In Chapter 8, we observe that relatively fair market income distribution may at least partly account for the East Asian puzzle. The low unemployment rates in East Asia seem to work as ‘magic bullets’ by reducing its welfare state’s burden, narrowing its market income inequality and the relieving the maleficent effects of ‘worklessness’ on the otherwise unemployed population. The findings suggest that fairer distribution in the market before redistribution through the welfare states may have more significant effects on aggregate health (in Chapters 7). This approach, emphasising the role of fairer markets rather that of welfare states, may have implications considering the current status of welfare states "between increasing demands and constrained resources" (Taylor-Gooby, 2001, p. 133).

Second, the mysteriously high mortality rates among the Scandinavian old population needs to be addressed from a policy perspective. Arguably, the Scandinavian welfare states have the dramatic age thresholds because they still have one of the best infant mortality rates among the welfare regimes just behind East Asian welfare regime, but its mortality rate among the aged over 65 is the worst among the welfare regimes as seen Figure 6-1. However, the general expectation regarding their longevity is still prevalent. A typical example is a recent article by a British media outlet, 'The Guardian'. In a special series on the NHS, the broadsheet newspaper acclaims the Danish social system for its elderly population.

"With most rich countries, including the UK, trying to work out how to look after growing numbers of old people using finite resources, Denmark may offer several solutions. The country spends 2.2% of its GDP on care for the elderly, second only to Sweden, and Danes over the age of 65 receive a basic pension of about 8,000 krone (£811) a month, before tax" (Russell, 2016).

According to the OECD (2015f), Denmark has the worst figures even among the struggling Scandinavian states, as Danish female life expectancy at 65 is 19.7 years in 2010, the second lowest after Czech Republic (19.0) among the 26 OECD nations selected in this thesis. The 26 nations' average is 21.4 years. Danish male life expectancy at 65 is also the second worst at 17.0 year, only higher than Czech Republic (15.5) again and one year below the 26 nations' average (18.0 year). Given this, Denmark cannot be an exemplary state especially in consideration of its huge investment in elderly health, as reported in the newspaper article.

To further analyse this puzzling findings in more detail, the Danish capital Copenhagen could be a good subject of a case study or a policy intervention because its old-age mortality rate (5.15 deaths per 100 people aged over 65 annually) is not only the highest among the Scandinavian regions but also the 22nd highest among the selected 292 regions. Copenhagen's statistics are the same as those of U.S. West Virginia and Turkey's Eastern Marmara, which borders the eastern side of Istanbul. Its infant mortality rate (4.1 deaths) is also the highest among the Scandinavian regions and the 105th highest among the 292 regions.

If Copenhagen is the representative case of the Scandinavian puzzle, Norway's rural Hedmark and Oppland region is the typical case of the contrasting pattern between positive infant mortality rate and negative elderly mortality record: the age threshold in the Scandinavian puzzle. Among the 292 regions included in this thesis, the Norwegian inland region has the 54th lowest infant mortality rate but the 273th lowest old-age mortality rate. Hedmark and Oppland is a dramatic example of the age threshold effect because the region's peculiar health determinants, which have positive influence on its relatively good infant mortality rate, may not have the same effect on the health of its elderly population. Given that the age threshold effects can be observed throughout the Scandinavian regions, the Norwegian inland region can be a target of a case study for future policy intervention. On the other hand, the surprisingly good health results in East Asia require closer scrutiny to find out what factors cause the unexpected outcomes.

Impressive health records are reported not only in Japan and South Korea but also in other East Asian nations (see Chapter 8).

Third, economic growth needs still to be a policy target to enhance aggregate health, but caution also needs to be exercised. The findings in Chapters 6 and 7 refute the presence of 'GDP per capita threshold effect' over which the economic growth does not have positive influence on aggregate health any more. We find that the economic growth even among the richest nations and regions can contribute to the enhancement of aggregate health. This finding is also compatible with the evidence supporting the 'broken limits to life expectancy' (Oeppen and Vaupel, 2002). Therefore, policies aiming at the economic growth need to be pursued for the sake of aggregate health.

However, externalities such as pollution or depletion of energy or goods also need to be taken into account in pursuit of the conventional beliefs in economic growth. Combining the externalities, commentators contend that economic growth may be degraded into 'de-growth' or 'uneconomic growth' (e.g. Baykan, 2007, Lawn, 2008) because economic growth that ignores the serious side effects does not only harm human wellbeing but also is not sustainable. It should be noted that air quality is found to be one of the strongest determinants of most of aggregate health indicators in Chapter 6.

Fourth and finally, the largely unidentified causes of the second Scandinavian and East Asian puzzles may have also significant implications for health policy. In this thesis, the two puzzles have been identified but their causes are not fully explained despite some suggestions forwarded in Chapter 8. As discussed, the hidden causes, which are favourable to East Asian health and detrimental to Scandinavian health, are powerful enough to eclipse the positive effects of general health determinants such as high incomes, narrow income inequalities, clean air and high educational attainment in Scandinavian regions. Without recognising the hidden causes or unobserved variables, we cannot only resolve the two puzzles but also miss the potential policy intervention to enhance the Scandinavian aggregate health.

In conclusion, this thesis, starting with the two research questions on the relationship between income inequality and aggregate health and welfare regimes and aggregate health, manages to find answers with some empirical evidence. On the first research question, which directly involves the controversial Wilkinson Hypothesis, this thesis finds that there might be four threshold effects over GDP per capita, age, gender and period. Over each threshold, the hypothesised relationship could reverse and vanish. First of all, this thesis presents statistical evidence to refute the GDP per capita threshold and the income indicator turns out to have statistically significant associations with most of health indicators over the threshold. Second, the findings in this thesis also demonstrate that Wilkinson Hypothesis works for some population groups such as infant, child or either genders but not for older generations. The period threshold, suggested in this thesis, is not tested here and awaits future studies. In the end, the findings in this thesis have mixed evidence for the Wilkinson Hypothesis with the hypothesised relationship often vanishing over the thresholds.

On the second research question - concerning the relationship between welfare regimes and aggregate health - the findings in this thesis confirm the presence of the second Scandinavian puzzle. Scandinavian welfare regimes surprisingly underperform in all of the tested aggregate health indicators. On the other hand, the East Asian welfare regime reports markedly good health outcomes despite its relatively low average income level and relatively wide disposable income inequality. The unexpected outcomes also confirm what this thesis proposes as the East Asian puzzle. Even though this thesis suggests several accounts for the paradoxical findings, the pair of puzzles poses serious challenges to conventional studies on determinants of aggregate health.

<APPENDICES>

Appendix 2-1) Theoretical Pathways And Variables Operationalised.

Theo-retical Pathways	Criticism		Implications for the thesis	Possible Variables	Possible variables in available regional/national datasets
1. Artefact	The term of 'artefact' is misleading as the pathway is related to aggregation of individual's health enhancement even in logical terms (Subramanian & Kawachi, 2004)		Income inequality can be still be a significant variable in influencing aggregate health (M. Marmot & Wilkinson, 2001)	① Income inequality	**In regional level:** Gini index (OECD regional dataset) **In national level:** Gini index, Theil index, 20/80 ratio
2. Social Selection	Social selection accounts requires acceptance of two disproved assumptions on genetic-racial-disease relationships (Goodman, 2000)		The account's contribution is little or negligible (Lundberg, 1991; Marmot et al, 1997).	Variables probably not needed for this thesis	No possible variables
3. Materialist	Those who support '1. artefact' account sceptic on other regional-	There is a threshold over which the income-health relationships no more work (Wilkinson & Pickett, 2009)	The need arises to test whether the threshold does exist or not with there being contrasting conclusions of its supporters (Nowatzki, 2012; Pop et al, 2014) and sceptics (Babones, 2008; Torre & Myrskuyla,	① Income inequality, ② income level, physical environment such as housing & transporation,	**In regional level:** GDP per capita, Gini index (OECD regional dataset) **In national level:** GDP per capita, Gini index, ratio of

	level determinants of aggregate health		2014).	③education, ④ occupation	attaining tertiary education among 25~64 (OECD national dataset)
4. Psycho-social	(Gravelle, 1998; Muller, 2002)	"The cultural meaning of economic inequality is also likely to vary and make a difference to outcomes" (Rowlingson, 2011; 26).	Just like the "3. materialists", contenders of this psychosocial pathways also use income inequality as the independent variable when it comes statistical analysis. There is no reason not to use the same variables as those for "3. Materialist pathways"	① Income inequality, ② income level, physical environment such as housing & transportation, ③ education, ④ occupation	**In regional level:** GDP per capita, Gini index (OECD regional dataset) **In national level:** GDP per capita, one of these inequality indexes (Gini index, Theil index, 20/80 ratio), ratio of attaining tertiary education among 25~64 (OECD national dataset)
5. Social Capital		"Strong evidence" suggests that people with a lower socioeconomic status generally have lower levels of social capital" (Uphoff, Pickett, Cabieses,	It is unclear that social capital is an independent factor or a merely dependent factor of individual's socioeconomic status. This thesis does not adopt the social capital variables.	Generally surveys on people's trust in other people or civic participation (Rose, 2000; Veenstra, 2002),	No possible variables

		Small, & Wright, 2013).		but not used in this thesis	
6. Behavioural/Cultural		Behaviours such as smoking and drinking can be outcomes or mediating factors rather than independent determinants of ill health themselves (Barnett, Moon, & Kearns, 2004; Pampel, 2002)	"Cultural characteristics and ingrained behaviours undoubtedly play a role also, at least in particular countries and should be included where relevant" (Starfield, 2007; 1360).	Health hazardous habits (⑤ alcohol ⑥ smoking) (Scarborough et al., 2011), ⑦ dietary characteristics (Prättälä et al., 2009)	**In regional level:** Not available **In national level:** alcohol consumption (60~), tobacco consumption (60~) - OECD national dataset
7. Environmental		Few empirical studies on this subject, and few criticisms on it.	"The quality of people's local environment, mainly water and air, has a direct impact on their collective health" (OECD, 2015a).	⑧ water quality, ⑨ air quality	**In regional level:** Air pollution index (level of PM2.5) **In national level:** not available

8. Policy Pathway		Medical care is only “the ambulance waiting at the bottom of the cliff” (Daniels, 2001: 6)	Health system (Conley & Springer, 2001) or welfare policy (Engster & Stensoeta, 2011) or even welfare regime (Bambra, 2006a) still have significant effect on aggregate health	⑩ public health, ⑪ welfare state characteristics, ⑫ welfare regime type	**In regional level:** redistribution index (tax and transfer), welfare regime type of the regions **In national level:** redistribution index (tax and transfer), public health spending, welfare regime type of the nations

Appendix 4-1) List of Primary Articles in Table 4-5

	Infant mortality rate		Child mortality rate		working age mortality		old age mortality		life expectancy		all-age mortality		life expectancy losses
Welfare regime	Bambra (2006),Chuang et al (2012),Chung & Muntaner (2007),Raphael (2013)	Karim et al (2010),Regidor et al (2011)						Popham, Dibben & Bambra (2013)		Chuang et al (2012),Kangas (2010), Karim et al (2010),Popham, Dibben & Bambra (2013),Raphael (2013)			
Income Inequality	Babones (2008),Beckfield (2004),Collison (2007),Lindström& Lindström(2006),Lynch et al (2001),Macinko, Shi & Starfield (2004),Muntaner et al (2002),Navarro et al (2003),Navarro et al (2006),Ram (2006),Regidor et al (2011),Wu et al (2008)	Avendano (2012),Beckfield (2004),Chung & Muntaner (2008),Leigh & Jencks (2007),Mello & Milyo (2001),Torre & Myrskylä (2014)	Lynch et al (2001),Muntaner et al (2002), Sengoelge (2013), Torre & Myrskylä (2014)		Ross et al (2005)	Lindström& Lindström(2006),Lynch et al (2001),Muntaner et al (2002),Torre & Myrskylä (2014)		Lynch et al (2001),Muntaner et al (2002),Torre & Myrskylä (2014)	Babones (2008),Beckfield (2004),De Vogli (2004),Ellison (2002),Ram (2006),Torre & Myrskylä (2014)	Beckfield (2004),Gravelle, Wildman & Sutton (2002),Leigh & Jencks (2007),Lindström& Lindström(2006),Lynch et al (2001),Mello & Milyo (2001),Muntaner et al (2002),Navarro et al (2003), Navarro et al (2006), Pop et al (2013)	Kondo et al (2009)	Lynch et al (2001),Muntaner et al (2002)	Shkolnikov et al (2011)
political tradition	Muntaner et al (2002),Navarro & Shi (2001)		Muntaner et al (2002)			Muntaner et al (2002)		Muntaner et al (2002)		Muntaner et al (2002)		Muntaner et al (2002)	
welfare state spending (generocity)	Engster & Stensota (2011),Ferrarini & Norstrom (2010),Lundberg et al (2008),Muntaner et al (2002),van der Heuvel et al (2013)	Wu et al (2008)	Muntaner et al (2002)			Muntaner et al (2002)	Lundberg et al (2008),Norstrom & Palme (2010)	Muntaner et al (2002)	Kangas (2010)	Muntaner et al (2002)		Muntaner et al (2002)	
public health spending or coverage	Chung & Muntaner (2008),Conley & Springer (2001),Macinko, Shi & Starfield (2004),Navarro et al (2003), Navarro et al (2006)	Regidor et al (2011),Wu et al (2008)								Navarro et al (2003),Navarro et al (2006)			
wealth inequality	Nowatzki (2012)								Nowatzki (2012)				
state by state comparison					Harding et al (2013)					Vaupel et al (2011)			

Appendix 4-2) List of Primary Articles in Table 4-6

	Infant mortality rate		Child mortality rate		working age mortality		old age mortality		life expectancy		all-age mortality		life expectancy losses
descrpitive	Navarro & Shi (2001),Raphael (2013),van der Heuvel et al (2013)				Harding et al (2013)					Kangas (2010),Raphael (2013)			
sationary analysis without control variables	Bambra (2006),Collison (2007),Navarro et al (2003),Navarro et al (2006)				Ross et al (2005)					Navarro et al (2003),Navarro et al (2006)			
stationary analysis with control variables	Babones (2008),Lindström & Lindström (2006),Lynch et al (2001),Muntaner et al (2002),Nowatzki (2012),Wu et al (2008)	Karim et al (2010), Wu et al (2008)	Lynch et al (2001),Muntaner et al (2002), Sengoelge (2013)			Lindström & Lindström (2006),Lynch et al (2001),Muntaner et al (2002)		Lynch et al (2001),Muntaner et al (2002)	Babones (2008),De Vogli (2004),Ellison (2002),Nowatzki (2012)	Karim et al (2010),Lindström & Lindström (2006),Lynch et al (2001),Muntaner et al (2002)		Lynch et al (2001),Muntaner et al (2002)	
multilevel analysis	Chuang et al (2012),Chung & Muntaner (2007)									Chuang et al (2012)			
decompositional analysis								Popham, Dibben & Bambra (2013)		Popham, Dibben & Bambra (2013),Vaupel et al (2011)			Shkolnikov et al (2011)
time-series regression	Beckfield (2004),Chung & Muntaner (2008),Engster & Stensota (2011),Regidor et al (2011)	Chung & Muntaner (2008),Regidor et al (2011)							Beckfield (2004),Kangas (2010)	Gravelle, Wildman & Sutton (2002)			
fixed or random effects model	Babones (2008),Chung & Muntaner (2007),Conley & Springer (2001),Ferrarini & Norstrom (2010),Lundberg et al (2008),Macinko, Shi & Starfield (2004),Ram (2006)	Avendano (2012),Beckfield (2004),Leigh & Jencks (2007),Mello & Milyo (2001),Torre & Myrskylä (2014)	Torre & Myrskylä (2014)			Torre & Myrskylä (2014)	Lundberg et al (2008),Norstrom & Palme (2010)	Torre & Myrskylä (2014)	Babones (2008),Ram (2006),Torre & Myrskylä (2014)	Beckfield (2004),Leigh & Jencks (2007),Mello & Milyo (2001),Pop et al (2013)	Kondo et al (2009)		

Appendix 4-3) List of Primary Articles in Table 4-7

	Infant mortality rate		Child mortality rate		working age mortality		old age mortality		life expectancy		all-age mortality	life expectancy losses
OECD	Bambra (2006),Chuang et al (2012),Chung & Muntaner (2007),Conley & Springer (2001),Engster & Stensota (2011),Macinko, Shi & Starfield (2004),Navarro & Shi (2001),Nowatzki (2012),Raphael (2013)	Avendano (2012)							Nowatzki (2012)	Chuang et al (2012),Raphael (2013)		
WHOSIS						Lindström& Lindström(2006)				Lynch et al (2001),Muntaner et al (2002)		
WHO Mortality	Lynch et al (2001),Muntaner et al (2002)		Lynch et al (2001),Muntaner et al (2002), Sengoelge (2013)			Lynch et al (2001), Muntaner et al (2002)		Lynch et al (2001),Muntaner et al (2002)			Lynch et al (2001),Muntaner et al (2002)	
Human Mortality DB	Ferrarini & Norstrom (2010)	Leigh & Jencks (2007),Torre & Myrskylä (2014)	Torre & Myrskylä (2014)			Torre & Myrskylä (2014)		Popham, Dibben & Bambra (2013),Torre & Myrskylä (2014)	Kangas (2010), Torre & Myrskylä (2014)	Kangas (2010), Leigh & Jencks (2007),Popham, Dibben & Bambra (2013),Vaupel et al (2011)		Shkolnikov et al (2011)
World Bank WDI	Babones (2008), Ram (2006)								Babones (2008)			
World Bank WDR									Ellison (2002),Ram (2006)			
individual nations DB					Harding et al (2013), Ross et al (2005)							
multiple DBs	Chung & Muntaner (2008),Lundberg et al (2008),Regidor et al (2011),van der Heuvel et al (2013),Wu et al (2008)	Chung & Muntaner (2008), Regidor et al (2011),Wu et al (2008)	Chung & Muntaner (2008)				Lundberg et al (2008),Norstrom & Palme (2010)					

Appendix 4-4) Summary of 48 Reviewed Primary Articles

author	year	data	study design/ methods/ target	No. of nations	health indicator	independent variable	conclusions
Conley and Springer	2001	1. comparative welfare data set, OECD health dataset, etc	6. country fixed effect model, pooled time-series with PCSE	19 (349 nation-years)	1. infant mortality rates	5. state health spending	pro-theory
Navarro and Shi	2001	1. OECD health data in 1998, comparative welfare state dataset	1. mainly descriptive	18 nations	1. infant mortality rates	3. political tradition	pro-theory
Lynch et al	2001	2, 3. Luxembourg Income Study, WHO Statistical Information system(for LE), WHO mortality database (for mortality)	3. Pearson correlation between income inequality and health outcomes weighted by population size, adjusted by income	16	1,2,3,4,5,6: age-specific mortality, life expectancy	2. income inequality, measures for social capital	mixed: 1,2 = pro-theory, 3,4,5,6 = anti-theory
Mello and Milyo	2001	9. Deininger and Squire (1996), Easterly (1999)	3, 6. pooled cross-section country-specific fixed model	12 nations (analysis on 30 nations include developing nations)	1,5: life expectancy, infant mortality	2: the Gini coefficient (explanatory variables: GDP per capita and secondary school enrolment)	anti-theory

Muntaner et al	2002	2,3: Luxembourg Income Study, World Values Survey, Comparative Welfare States Dataset, World Health Organization's Statistical Information System, WHO Mortality Database	3. Pearson correlation between social capital, economic inequality, working-class power and population health indicator weighted by population size	16 wealthy nations	1,2,3,4,5,6: life expectancy, SRH, low birth weight, age-specific mortality	2,3,4: economic inequality, working-class power, social capital, welfare state spending	mixed: 1,2 = pro-theory, 3,4,5,6 = anti-theory
Gravelle, Wildman and Sutton	2002	9. the Deininger and Squire (1996) World Tables, Penn World Tables Mark 5, US census international database	3. regression analysis after Box Cox transformation of variables after adjusting for income	high income nations over $5000 (The number of nations not specified)	5. male life expectancy	2: Gini, GDP per capita	anti-theory
Ellison	2002	6. World Bank World Development Report	3. multiple Log10-linear regressions of life expectancy against GDP per capita with income inequality (to test relative income hypothesis in relation the 'artefact' contention)	22 high-income nations	5. life expectancy	2. GDP per capita, income inequality measures	pro-theory

Navarro et al	2003	9. Comparative welfare data set, Navarro, Schmitt and Astudillo (2003), health data sources not specified	6. pooled cross-sectional study,	17 OECD nations	1,5. infant mortality rate, life expectancy	2,5. labour market, welfare state (*in fact, public health spending), income inequality	mixed: 1 = pro-theory, 5 = anti-theory
Coburn	2004	1. OECD health data, Gottschalk and Smeeding (2000), WHO Statistics Annual	3. compare infant mortality rate between welfare regimes/ analyse correlation between decommodification and health indicators	18 nations	1,7. age-adjusted mortality rate, infant mortality rate, potential years of life lost (PYLL)	1. welfare regime, decommodification,	*generally pro-theory, but cannot be coded as p-value not given
De Vogli	2005	9. United Nations Development Program's human development indicators database (2003)	3. The correlation weighted by population size and adjusted for per capita gross domestic product	21 nations (Japan)	5. life expectancy	2. income inequality	pro-theory
Beckfield	2004	9. the World Bank's World Tables, the UN's World Population Prospects	3,6. OLS models to assess whether income inequality affects health net of technical controls, economic development, and time trends. Also fixed-effects models to account for unmeasured heterogeneity that can bias OLS estimates	rich nation with GDP over $ 5000, 6,000, 7000, 8000 (the number not specified)	1,5. life expectancy, infant mortality	2. income inequality (Real GDP Per Capita and Year of Observation as control variable)	mixed: pro-theory theory for regression, anti-theory for fixed model

Macinko, Shi and Starfield	2004	1. World Value Surveys, Luxembourg Income Study, OECD Health Data 2001 database	6. fixed-effects multivariate regression, pooled, cross-sectional, time-series study	19 OECD member nations	1. infant mortality rate	2,5. Theil income inequality index, GDP per capita, health expenditure, unemployment, etc	pro-theory
Ross et al	2005	7.from each nations, for example, Australian data are from the 1991 Census and Australian Bureau of Statistics death registrations	2. bivariate linear regression analyses and weighted analysis of variance (ANCOVA)	528 metropolitan areas in five nations (Australia, Canada, UK, Sweden, US)	3. working-age (25–64) mortality	2. income inequality (as measured by median share of income)	pro-theory
Navarro et al	2006	9. Müller and Strom (2000), ILO-LABORSTA, health datasets not specified	2. examine the interactions between political traditions, policies, and public health outcomes by testing correlations	17 European nations	1,5. infant mortality rate, life expectancy	2,5. four variables on power relations, 4 on labour market, 2 on welfare state and 2 on economic inequality	pro-theory
Lindström and Lindström	2006	2,9. World Bank Group, WHO online database, Inglehart et al (1998)	3. multivariate linear regression model	11 rich nation with over 12500 GDP per capita	1,3,5. adult mortality rate (25–64years), life expectancy, and infant mortality rate	2. social capital, and GNP (gross national product) per capita and Gini index	mixed: 1 = pro-theory, 3,5= anti-theory

Bambra	2006	1. OECD Health Database	2. compare the average IMR between three regimes and test correlations between infant mortality rate and decommodification	18 nations	1. infant mortality rate	1. decommodification score, welfare state regime	pro-theory
Leigh and Jencks	2007	4. Human Mortality Database, Leigh (2006)	6. country fixed or year fixed effects of income inequality.	12 nations	1,5. life expectancy, infant mortality rate	2. the share of pre-tax income going to the richest 10% of the population	anti-theory
Collison	2007	9. UNICEF's 'State of the World's Children' reports 2003–2006	2. Pearson correlation coefficients	21 OECD members (Korea, Japan)	1. under-five mortality rate	2. income inequality ratio	pro-theory
Chung and Muntaner	2007	1. OECD helath data 2000, UN Common Statistical Database	4,6. two-level multilevel model, fixed effects of welfare state tested	19 OECD nations	1. IMR	1. GDP per capita, welfare regime	pro-theory
Lundberg et al	2008	8. the Social Citizenship Indicator Program, the Human Mortality Database and from the WHO Mortality Database.	6. pooled cross-sectional time-series analyses with PCSE	18 OECD nations, 70-00	1,4: IMR, old-age excess mortality rate	4. family policy generosity, pension policy generosity	pro-theory
Chung and Muntaner	2008	8. OECD, UNICEF	6. panel data analysis using the robust cluster variance estimator	19 nations	1. LBW, IMR, U5MR	2, 5. economic, political, welfare state variables	mixed: income inequality = anti-theory, public health= pro-theory

Babones	2008	5. World Institute for Development Economics Research (2000) for 1970-1995, World Bank (2004) World development indicators database.	3, 6. simple correlation, fixed effects model	30 developed nations	1, 5. life expectancy, IMR	2. Gini	pro-theory
Wu et al	2008	8. UNICEF's The State of the World's Children 2004, United Nations' Demographic Yearbook for IMR, The World Health Report 2004 for the U5MR	2. multiple regression model for mortality, GDP, national health spending, Gini, etc.	21 nations (Taiwan, Japan)	1. IMR, Under-five mortality rate	2,4,5. GDP, national health spending, Gini,	mixed.
Kondo et al	2009	9. Peer reviewed papers with multilevel data from PubMed, the ISI Web of Science, and the National Bureau for Economic Research database.	6. Random effects meta-analyses, calculating the overall relative risk for subsequent mortality among prospective cohort studies and the overall odds ratio	meta anlaysis	6. SRH, mortality	2. Gini	pro-theory.
Kangas	2010	4. Human Mortality Database 1900-2000, Maddison (2003)	1,6. Pooled cross-sectional times series analysis, Prais-Winsten regressions on correlated panels and corrected standard errors (PCTS)	17 OECD nations	5. life expectancy	1,4. welfare state spending, welfare regime	mixed: 1 = anti-theory, 4= pro-theory

Ferrarini and Norstrom	2010	4. Human Mortality Database (2006), Social Citizenship Indicator Program (SCIP, 2008), the Parental Leave Benefit Database (PAL, 2009)	6. time series analysis, fixed effect modelling with PCSE	18 welfare democracies, 70-00	1. IMR	4. family policy legislation, economic development	pro-theory
Norstrom and Palme	2010	8. WHO Mortality Data Base, Human Mortality Database, data from Social Citizenship Indicator Program	6. pooled cross-sectional time-series analyses, fixed effects model with PCSE	18 OECD nations	4. old-age excess mortality (ratio of mortality 65+ to mortality 30–59)	4. pension rights: basic security and income security	pro-theory
Karim et al	2010	9. The World Factbook 2003	3. one-way ANOVA tests and one-way ANCOVA tests to test between-group variances	30 nations	1,5. IMR, life expectancy	1. welfare regime	anti-theory
Regidor et al	2011	8. OECD Health Data 1960~, Human Mortality Data Base	6. generalized estimating equations (GEE)	17 Western nations, 80-05	1. IMR	1,2,5. political tradition, family policy model, public health expenditure, income inequality	mixed: By the early 21st century, the differences (between regimes) in infant mortality were negligible. The relationship between public health expenditure and infant mortality disappears beginning in 1995. increased income inequality is associated with higher infant mortality.
Engster and Stensota	2011	1. panel data from 20 OECD nations (95, 00, 05)	6. multivariate regression analysis (OLS) with robust-cluster variance estimator	20 OECD nations (no Japan)	1. under 5 year mortality rate	4. family policy generosity (GDP per capita, demographic variables as control variables)	pro-theory

Vaupel et al	2011	4. Human Mortality Database (1840-2009)	5. the original focus is to analyse the relations between aggregate health and health inequalities	40 nations	5. life expectancy	7. state by state comparison	anti-theory
Shkolnikov	2011	4. World Income Inequality Database (2008), Human Mortality Database (2007), OECD statistics (2008)	6. country and time fixed effects regressions as Gini and time dummy variables as independent variables. since 1975	17 nations (Japan)	7. the average life expectancy losses caused by death at age	2. Gini index, time dummy variables	neutral, coding as anti-theory:
Chuang et al	2012	1. OECD Dataset, World Development Indicators, Asia Development Bank (80~06)	4. multilevel random intercept model	31 nations (East Asia)	1,5. IMR, life expectancy	1. welfare regime	mixed: 1 - pro-theory, 5 - anti-theory
Nowatzki	2012	1. Davies et al (2007), OECD health data (2009)	3. bivariate cross-sectional analyses of the relationship between wealth inequality (Gini coefficient) and population health	14 wealthy nations (Japan)	1,5. life expectancy, IMR	6. wealth inequality	pro-theory
Avendano	2012	1. Standardized World Income Inequality Database, OECD health data	6. country and year fixed effects model to capture evolution of income inequality across countries 60-08	34 OECD nations, 60-08	1. IMR	2. income inequality	anti-theory
Harding et al	2013	7: UK Office for National Statistics Longitudinal Study (ONS-LS), Turin Longitudinal Study (TLS) and the Finnish linked register study (FS).	1: Death rates and rate ratios (RRs) (reference rates = ‘in-work’), 1970 s–2000 s, were estimated for those aged 45–64 years and compared	Three nations (Italy, Finland, England and Wales)	3. mortality, death rate ratio	7. state by state comparison. whether in-work or not-in-work	Death rates for the not-in-work were lowest in Turin and highest in Finland, but opposite for in-works

Raphael	2013	1. OECD Health at a Glance (2009), OECD Social Expenditure Database (2008)	1. comparison of health indicators between regimes/ecological, descriptive	21 Western OECD members	1,5. life expectancy, infant mortality	1. key social determinants (income inequality, family poverty, union density, collective agreement), public commitments through expenditures	mixed: 1 = pro-theory, 5 = anti-theory
Popham, Dibben and Bambra	2013	4. Human Mortality Database	5. measuring life expectancy and life expectancy lost per death	37 nations (Japan, Taiwan)	4,5. average life expectancy lost per death	1. welfare regime, age, sex	**(Men)** Nordic countries on average lost least life expectancy and gained most equality due to having low premature mortality **(Women)** Nordic countries were not the most equal on average nor did they have the highest life expectancy
van der Heuvel et al	2013	8. OECD, World Bank, and UNICEF	1. descriptively compare key indicators among the five nations	Five nations (Sweden, the Netherlands, Canada, U.S., Cuba)	1. IMR, LBW, Under five mortality	4. redistributive policies on prenatal care, 2) maternal leave, 3) child health care, and 4) child care and early childhood education	pro-theory

Sengoelge	2013	3. European Union Income Social Inclusion and Living Conditions Database, WHO mortality database	3. Linear regression analysis to see relationship between CM and GDP, II mediated by housing/neighbourhood condition, ecological / 1-14 aged	26 European	2. child mortality	2. II, GDP, housing, neighbourhood conditions	pro-theory
Pop et al	2013	9. the Standardized World Income Inequality Database (SWIID), United Nations' World Prospects (2009).	6. hybrid fixed effects regression	29 rich countries (591 country-year)	5. life expectancy	2. Gini, GDP per capita	anti-theory
Torre and Myrskyläa	2014	4. human mortality database (2010)	6. country fixed effect model and quinquinnial time series to analyse II and variation in age at death	21 developed nations, 75-06	1,2,3,4,5. age-specific mortality	2. Gini, GDP per capita	mixed: 2,5 = pro-theory, 1,3,4 = anti-theory
Nelson and Fritzell	2014	1. OECD health data	6. fixed effects pooled time-series regression with cluster robust standard error	18 countries 90-09	5,6. age-standardised mortality, LE	4. minimum income benefits, GDP	pro-theory
Rambotti	2015	9. UN Human Development Report	3. cross-national multiple regression	20 nations	5. LE	2. income inequality, relative poverty	5. anti-theory
Safaei	2015	1. OECD health data	3. cross-national multiple regression	31 OECD members	1,2,5. IMR, CMR, LE	8. pro-primary distributional stance, pro-secondary stance	5=anti-theory, mixed for IMR, CMR
Bocoum, Macombe and Reveret	2015	8. UNU-WIDER, World Bank	6. generalized method of moments (GMM)	29 OECD nations	1. IMR	2. Gini, GDP	1. pro-theory

Fritzell et al	2015	3. WHO Mortality Database	6. CSTS method with panel corrected standard errors	30 countries 78-10	1,2,3. mortality of infant, child and adult	1. poverty rate, regime type	mixed: 1=pro-theory, 2,3= anti-theory
Hu et al	2015	8. Human Lifetable Database etc.	6. country and time fixed effects pooled cross-section regression	43 European nations 87-08	1.5. LE, IMR	2. Gini, GDP	1.5. anti -theory

<REFERENCES>

Abrahamson, P. (1999). The Scandinavian model of welfare. In Abrarahamson, P., Bouget, D., Palier, B. (Ed.) *Comparing social welfare systems in Nordic Europe and France*. (pp. 31-60), Paris: Service de l'Information et de la Communication. Retrieved from http://www.researchgate.net/profile/Peter_Abrahamson/publication/28367839_The_Scandinavian_model_of_welfare/links/00b7d51516c0093cc6000000.pdf

Abrahamson, P., and Wehner, C. (2006). Family and/or Work in Europe? *Journal of Comparative Family Studies*, *37*(2), 153–171.

Achen, C. H. (2000). Why lagged dependent variables can suppress the explanatory power of other independent variables. Political Methodology Working Paper. *Retrieved from http://www-personal.umich.edu/~franzese/Achen.2000.LDVstealingExplanPower.pdf*

Acheson, D. (Ed.). (1998). *Independent inquiry into inequalities in health: report*. London: The Stationery Office.

Adamsson, V., Reumark, A., Fredriksson, I.-B., Hammarström, E., Vessby, B., Johansson, G., and Risérus, U. (2011). Effects of a healthy Nordic diet on cardiovascular risk factors in hypercholesterolaemic subjects: a randomized controlled trial (NORDIET): Diet and cardiovascular risk factors. *Journal of Internal Medicine*, *269*(2), 150–159.

Adler, J., and Parmryd, I. (2010). Quantifying colocalization by correlation: The Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytometry Part A*, *77A*(8), 733–742.

Aidukaite, J. (2011). Welfare reforms and socio-economic trends in the 10 new EU member states of Central and Eastern Europe. *Communist and Post-Communist Studies*, *44*(3), 211–219.

Anderson, T., and Kohler, H.-P. (2013). Education Fever and the East Asian Fertility Puzzle. *Asian Population Studies*, *9*(2), 196–215.

Armstrong, B., and Doll, R. (1975). Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices. *International Journal of Cancer*, *15*(4), 617–631.

Aspalter, C. (2006). The East Asian welfare model. *International Journal of Social Welfare*, *15*(4), 290–301.

Asthana, S., and Halliday, J. (2006). *What works in tackling health inequalities?: pathways, policies and practice through the lifecourse*. Bristol: Policy Press.

Atkinson, A. B., & Brandolini, A. (2009). On data: a case study of the evolution of income inequality across time and across countries. *Cambridge Journal of Economics*, *33*(3), 381-404.

Auger, N., Le Serbon, E., and Rostila, M. (2015). Leaving Sweden behind: Gains in life expectancy in Canada. *Scandinavian Journal of Public Health*, *43*(4), 340–347.

Avendano, M., and Hessel, P. (2015). The income inequality hypothesis rejected? *European Journal of Epidemiology*, *30*(8), 595–598.

Babones, S. J. (2008). Income inequality and population health: Correlation and causality. *Social Science & Medicine*, *66*(7), 1614–1626.

Ballas, D., Dorling, D., Nakaya, T., Tunstall, H., and Hanaoka, K. (2014). Income inequalities in Japan and the UK: a comparative study of two island economies. *Social Policy and Society*, *13*(1), 103–117.

Bambra, C. (2004). The worlds of welfare: illusory and gender blind? *Social Policy and Society*, *3*(3), 201–211.

Bambra, C. (2005). Cash Versus Services: and the Decommodification of Cash Benefits and Health Care Services. *Journal of Social Policy*, *34*(2), 195–213.

Bambra, C. (2006a). Health Status and the Worlds of Welfare. *Social Policy and Society*, *5*(1), 53-62.

Bambra, C. (2006b). Research Note: Decommodification and the worlds of welfare revisited. *Journal of European Social Policy*, *16*(1), 73–80.

Bambra, C. (2011). Health inequalities and welfare state regimes: theoretical insights on a public health "puzzle." *Journal of Epidemiology and Community Health*, *65*(9), 740–745.

Bambra, C. (2011). Work, worklessness and the political economy of health inequalities. *Journal of Epidemiology and Community Health*, *65*(9), 746–750.

Bambra, C. (2013). In defence of (social) democracy: on health inequalities and the welfare state. *Journal of Epidemiology & Community Health*, *67*(9), 713–714.

Bambra, C., and Eikemo, T. A. (2009). Welfare state regimes, unemployment and health: a comparative study of the relationship between unemployment and self-reported health in 23 European countries. *Journal of Epidemiology and Community Health*, *63*(2), 92–98.

Barnett, R., Moon, G., and Kearns, R. (2004). Social inequality and ethnic differences in smoking in New Zealand. *Social Science & Medicine*, *59*(1), 129–143.

Bartley, M., Sacker, A., and Clarke, P. (2004). Employment status, employment conditions, and limiting illness: prospective evidence from the British household panel survey 1991–2001. *Journal of Epidemiology and Community Health*, *58*(6), 501–506.

Beck, N. (2001). Time-series-cross-section data: What have we learned in the past few years? *Annual Review of Political Science*, *4*(1), 271–293.

Beck, N., and Katz, J. N. (1995). What to do (and not to do) with time-series cross-section data. *American Political Science Review*, *89*(3), 634–647.

Beck, N., and Katz, J. N. (2004). Time-series-cross-section issues: dynamics, 2004. In *annual meeting of the Society for Political Methodology, Stanford University*. Retrieved from http://www.researchgate.net/profile/Jonathan_Katz5/publication/228723029_Time-series-cross-section_issues_dynamics_2004/links/0912f50cfa0073bdcb000000.pdf

Becker, L., and Oxman, A. (2008). Overviews of reviews. In Higgins, J., Green, S., (Ed.) *Cochrane Handbook for Systematic Reviews of Interventions*. (pp. 607-633). Chichester, England: Wiley.

Beckfield, J. (2004). Does Income Inequality Harm Health? New Cross-National Evidence. *Journal of Health and Social Behavior*, *45*(3), 231–248.

Benova, L., Cumming, O., and Campbell, O. M. R. (2014). Systematic review and meta-analysis: association between water and sanitation environment and maternal mortality. *Tropical Medicine & International Health*, *19*(4), 368–387.

Bergqvist, K., Yngwe, M. Åberg, and Lundberg, O. (2013). Understanding the role of welfare state characteristics for health and inequalities-an analytical review. *BMC Public Health*, *13*(1), 1234.

Bjørnskov, C., Dreher, A., Fischer, J. A. V., Schnellenbach, J., and Gehring, K. (2013). Inequality and happiness: When perceived social mobility and economic reality do not match. *Journal of Economic Behavior & Organization*, *91*, 75–92.

Black, N. (2001). Evidence based policy: proceed with care. *BMJ: British Medical Journal*, *323*(7307), 275.

Blakely, T. A., Kennedy, B. P., Glass, R., and Kawachi, I. (2000). What is the lag time between income inequality and health status? *Journal of Epidemiology and Community Health*, *54*(4), 318–319.

Blakely, T., Atkinson, J., and O'Dea, D. (2003). No association of income inequality with adult mortality within New Zealand: a multi-level study of 1.4 million 25–64 year olds. *Journal of Epidemiology and Community Health*, *57*(4), 279–284.

Blane, D., Smith, G. D., and Bartley, M. (1993). Social selection: what does it contribute to social class differences in health? *Sociology of Health & Illness*, *15*(1), 1–15.

Bloor, M., Samphier, M., and Prior, L. (1987). Artefact explanations of inequalities in health: an assessment of the evidence. *Sociology of Health & Illness*, *9*(3), 231–264.

Bobak, M., Pikhart, H., Rose, R., Hertzman, C., and Marmot, M. (2000). Socioeconomic factors, material inequalities and perceived control in self-rated health: Cross-sectional data from seven post-communist countries. *Social Science & Medicine*, *51*(9), 1343–1350.

Bollen, K. A., and Jackman, R. W. (1985). Regression Diagnostics An Expository Treatment of Outliers and Influential Cases. *Sociological Methods & Research*, *13*(4), 510–542.

Bonoli, G. (1997). Classifying Welfare States: a Two-dimension Approach. *Journal of Social Policy*, *26*(3), 351–372.

Borrell, C., Espelt, A., Rodríguez-Sanz, M., Burström, B., Muntaner, C., Pasarín, M. I., … Navarro, V. (2009). Analyzing Differences in the Magnitude of Socioeconomic Inequalities in Self-Perceived Health by Countries of Different Political Tradition in Europe. *International Journal of Health Services*, *39*(2), 321–341.

Brennenstuhl, S., Quesnel-Vallée, A., and McDonough, P. (2011). Welfare regimes, population health and health inequalities: a research synthesis. *Journal of Epidemiology and Community Health*, *66*(5), 397–409.

Breusch, T. S., and Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, *47*(5), 1287–1294.

Buğra, A., and Keyder, Ç. (2006). The Turkish welfare regime in transformation. *Journal of European Social Policy*, *16*(3), 211–228.

Cardinal, R. and Aitken, M. (2013). *ANOVA for the Behavioral Sciences Researcher* (1 edition). Mahwah, N.J: Psychology Press.

Carmines, E. G., McIver, J. P., and others. (1981). Analyzing models with unobserved variables: Analysis of covariance structures. *Social Measurement: Current Issues*, 65–115.

Cattan, M., White, M., Bond, J., and Learmouth, A. (2005). Preventing social isolation and loneliness among older people: a systematic review of health promotion interventions. *Ageing and Society*, *25*(1), 41–67.

Cattell, V. (2001). Poor people, poor places, and poor health: the mediating role of social networks and social capital. *Social Science & Medicine*, *52*(10), 1501–1516.

Chandola, T., Bartley, M., Sacker, A., Jenkinson, C., and Marmot, M. (2003). Health selection in the Whitehall II study, UK. *Social Science & Medicine*, *56*(10), 2059–2072.

Chandra, A., Martinez, G. M., Mosher, W. D., Abma, J. C., and Jones, J. (2005). Fertility, family planning, and reproductive health of US women: data from the 2002 National Survey of Family Growth. *Vital and Health Statistics. Series 23, Data from the National Survey of Family Growth*, (25), 1–160.

Chuang, Y.-C., Chuang, K.-Y., Chen, Y.-R., Shi, B.-W., and Yang, T.-H. (2012). Welfare state regimes, infant mortality and life expectancy: integrating evidence from East Asia. *Journal of Epidemiology and Community Health*, *66*(7), e23–e26.

Chuang, Y.-C., Cubbin, C., Ahn, D., and Winkleby, M. A. (2005). Effects of neighbourhood socioeconomic status and convenience store concentration on individual level smoking. *Journal of Epidemiology and Community Health*, *59*(7), 568–573.

Chung, H., and Muntaner, C. (2007). Welfare state matters: A typological multilevel analysis of wealthy countries. *Health Policy*, *80*(2), 328–339.

Chung, H., and Muntaner, C. (2008). Political and Welfare State Determinants of Infant and Children's Health Indicators: An Analysis of Wealthy Countries. *Gerencia Y Políticas de Salud*, *7*(14), 14-31.

CIA (2008) The World Factbook 2003. Retrieved 18th June, 2008. Retrieved by Karim et al (2010) from http://www.cia.gov/library/publications/download/download-2003/index.html.

Clarke, K. A. (2005). The Phantom Menace: Omitted Variable Bias in Econometric Research. *Conflict Management and Peace Science*, *22*(4), 341–352.

Coburn, D. (2004). Beyond the income inequality hypothesis: class, neo-liberalism, and health inequalities. *Social Science & Medicine*, *58*(1), 41–56.

Conley, D., and Springer, K. W. (2001). Welfare State and Infant Mortality1. *American Journal of Sociology*, *107*(3), 768–807.

Corneo, G. (2011). *Income Inequality, Value Systems, and Macro Economic Performance.* (GINI Discussion Paper No. 17). Amsterdam: AIAS. Retrieved from http://www.uva-aias.net/uploaded_files/publications/DP17-Corneo.pdf

Cousineau, D., and Chartier, S. (2010). Outliers detection and treatment: a review. *International Journal of Psychological Research*, *3*(1), 58–67.

Crimmins, E., Preston, S., and Cohen, B. (2011). *Explaining Divergent Levels of Longevity in High-income Countries*. Washington D.C.: National Academies Press.

Croissant, Y., Millo, G., and others. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software*, *27*(2), 1–43.

Dahl, E., Fritzell, J., Lahelma, E., Martikainen, P., Kunst, A., and Mackenbach, J. P. (2006). Welfare state regimes and health inequalities. In Siegrist, J. and Marmot, M. (ed) *Social Inequalities in Health: New Evidence and policy implications*. (pp. 193-222), Oxford: Oxford University Press.

Dahlgren, G., and Whitehead, M. (1991). *Policies and Strategies to Promote Social Equity in Health. Equity in Health*. Stockholm: Institute for Future Studies.

Daniels, N. (2001). Justice, health, and healthcare. *The American Journal of Bioethics: AJOB*, *1*(2), 2–16.

Daniels, N. (2007). *Just Health: Meeting Health Needs Fairly*. Cambridge; New York: Cambridge University Press.

Deaton, A., and Lubotsky, D. (2003). Mortality, inequality and race in American cities and states. *Social Science & Medicine*, *56*(6), 1139–1153.

De Vogli, R., Mistry, R., Gnesotto, R., and Cornia, G. A. (2005). Has the relation between income inequality and life expectancy disappeared? Evidence from Italy and top industrialised countries. *Journal of Epidemiology and Community Health*, *59*(2), 158–162.

Deeming, C., and Hayes, D. (2012). Worlds of Welfare Capitalism and Wellbeing: A Multilevel Analysis. *Journal of Social Policy*, *41*, 811–829.

Department of Health and Social Security. (1980). *Black Report.* London, U.K.: Department of Health and Social Security.

Dibben, C., Sigala, M., and Macfarlane, A. (2006). Area deprivation, individual factors and low birth weight in England: is there evidence of an "area effect"? *Journal of Epidemiology and Community Health*, *60*(12), 1053–1059.

Dion, M. (2006). Women's Welfare and Social Security Privatization in Mexico. *Social Politics: International Studies in Gender, State & Society*, *13*(3), 400–426.

Doorslaer, E. van, and Koolman, X. (2004). Explaining the differences in income-related health inequalities across European countries. *Health Economics*, *13*(7), 609–628.

Doorslaer, E. van, Wagstaff, A., and Bleichrodt, H. (1997). Income-related inequalities in health: some international comparisons. *Journal of Health Economics*, *16*(1), 93–112.

Dorling, D., and Barford, A. (2009). The inequality hypothesis: Thesis, antithesis, and a synthesis? *Health & Place*, *15*(4), 1166–1169.

Dougherty, C. (2011). *Introduction to Econometrics* (4 edition). Oxford ; New York: Oxford University Press.

Dragano, N., Siegrist, J., and Wahrendorf, M. (2010). Welfare regimes, labour policies and unhealthy psychosocial working conditions: a comparative study with 9917 older employees from 12 European countries. *Journal of Epidemiology & Community Health*, *65*(9), 793–799.

Drain, P. K., Smith, J. S., Hughes, J. P., Halperin, D. T., and Holmes, K. K. (2004). Correlates of national HIV seroprevalence: an ecologic analysis of 122 developing countries. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *35*(4), 407–420.

Duncan, C., Jones, K., and Moon, G. (1993). Do places matter? A multi-level analysis of regional variations in health-related behaviour in Britain. *Social Science & Medicine*, *37*(6), 725–733.

Ebbinghaus, B. (2005). When Less is More: Selection Problems in Large- N and Small-N Cross-National Comparisons. *International Sociology*, *20*(2), 133–152.

Ebbinghaus, B. (2012). Comparing Welfare State Regimes: Are Typologies an Ideal or Realistic Strategy?. Draft Paper presented at European Social Policy Network, *Edinburgh, UK*. Retrieved from http://www.espanet2012.info/__data/assets/pdf_file/0005/89033/Ebbinghaus_-_Stream_2.pdf

Eikemo, T. A., Bambra, C., Joyce, K., and Dahl, E. (2008a). Welfare state regimes and income-related health inequalities: a comparison of 23 European countries. *The European Journal of Public Health*, *18*(6), 593–599.

Eikemo, T. A., Bambra, C., Judge, K., and Ringdal, K. (2008b). Welfare state regimes and differences in self-perceived health in Europe: a multilevel analysis. *Social Science & Medicine (1982)*, *66*(11), 2281–2295.

Eikemo, T. A., Huisman, M., Bambra, C., and Kunst, A. E. (2008c). Health inequalities according to educational level in different welfare regimes: a comparison of 23 European countries. *Sociology of Health & Illness*, *30*(4), 565–582.

Ellaway, A., Macintyre, S., and Kearns, A. (2001). Perceptions of Place and Health in Socially Contrasting Neighbourhoods. *Urban Studies*, *38*(12), 2299–2316.

Emmenegger, P., Kvist, J., Marx, P., and Petersen, K. (2015). Three Worlds of Welfare Capitalism: The making of a classic. *Journal of European Social Policy*, *25*(1), 3–13.

Engster, D., and Stensoeta, H. O. (2011). Do Family Policy Regimes Matter for Childrens Well-Being? *Social Politics*, *18*(1), 82–124.

Esping-Andersen, G. (1990). *The three worlds of welfare capitalism*. Princeton, NJ: Princeton University Press.

Esping-Andersen, G. (1997). Hybrid or Unique?: the Japanese Welfare State Between Europe and America. *Journal of European Social Policy*, *7*(3), 179–189.

Esping-Andersen, G. (1999a). *Social Foundations of Postindustrial Economies*. Oxford, U.K.: Oxford University Press.

Esping-Andersen, G. (1999b). Preface for Chinese version. In *The Three Worlds of Welfare Capitalism*, Taipei: Mainstream Publisher.

Esping-Andersen, G. (2007). Multiple Regression in Small-N Comparisons. *Comparative Social Research*, *24*, 335–342.

Esping-Andersen, G. (2015). Welfare regimes and social stratification. *Journal of European Social Policy*, *25*(1), 124–134.

European Commission. (n.d.). *Efficiency and Effectiveness of Public Expenditure on Tertiary Education in the EU, Annex: Country Fiche Luxembourg* (European Economy Occasional Papers No. 70). Retrieved from http://ec.europa.eu/economy_finance/publications/occasional_paper/pdf/country_fiches/luxembourg.pdf

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., and Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *The FASEB Journal*, *22*(2), 338–342.

Fayissa, B. (2001). The determinants of infant and child mortality in developing countries: The case of Sub-Sahara Africa. *The Review of Black Political Economy*, *29*(2), 83–98.

Feng, Z., Wang, W. W., Jones, K., and Li, Y. (2012). An exploratory multilevel analysis of income, income inequality and self-rated health of the elderly in China. *Social Science & Medicine*, *75*(12), 2481–2492.

Fenger, H. J. M. (2007). Welfare Regimes in Central and Eastern Europe: Incorporating Post-Communist Countries in a Welfare Regime Typology. *Contemporary Issues and Ideas in Social Sciences*, *3*(2). Retrieved from http://journal.ciiss.net/index.php/ciiss/article/view/45

Fernquist, R. M. (2003). Perceived income inequality and suicide rates in Central/Eastern European countries and Western countries, 1990-1993. *Death Studies*, *27*(1), 63–80.

Ferragina, E., Seeleib-Kaiser, M., & Spreckelsen, T. (2015). The Four Worlds of 'Welfare Reality'–Social Risks and Outcomes in Europe. *Social Policy and Society*, *14*(02), 287-307.

Ferragina, E., Seeleib-Kaiser, M., and Tomlinson, M. (2013). Unemployment Protection and Family Policy at the Turn of the 21st Century: A Dynamic Approach to Welfare Regime Theory. *Social Policy & Administration*, *47*(7), 783–805.

Ferrarini, T., and Norström, T. (2010). Family policy, economic development and infant mortality: a longitudinal comparative analysis. *International Journal of Social Welfare*, *19*, S89–S102.

Ferreira, F. H., Lustig, N., & Teles, D. (2015). Appraising cross-national income inequality databases: An introduction. *The Journal of Economic Inequality*, *13*(4), 497-526.

Ferrera, M. (1996) 'The "Southern Model" of welfare in social Europe', *Journal of European Social Policy*,6(1), 17–37.

Fiscella, K., and Franks, P. (1997). Poverty or income inequality as predictor of mortality: longitudinal cohort study. *BMJ*, *314*(7096), 1724.

Ford, E. S., Giles, W. H., and Dietz, W. H. (2002). Prevalence of the metabolic syndrome among US adults: findings from the third National Health and Nutrition Examination Survey. *Jama*, *287*(3), 356–359.

Fouarge, D., and Layte, R. (2005). Welfare Regimes and Poverty Dynamics: The Duration and Recurrence of Poverty Spells in Europe. *Journal of Social Policy*, *34*(3), 407–426.

Franks, P., and Fiscella, K. (1998). Primary care physicians and specialists as personal physicians. Health care expenditures and mortality experience. *The Journal of Family Practice*, *47*(2), 105–109.

Frie, K. G., Eikemo, T. A., and Knesebeck, O. von dem. (2010). Education and self-reported health care seeking behaviour in European welfare regimes: results from the European Social Survey. *International Journal of Public Health*, *55*(3), 217–220.

Fritzell, J., Bäckman, O., and Ritakallio, V.-M. (2012). Income inequality and poverty: do the Nordic countries still constitute a family of their own? In J. Kvist, J. Fritzell, B. Hvinden, and O. Kangas (Eds.), *Changing social equality The Nordic welfare model in the 21st century* (pp. 165–186). Bristol, UK: Policy Press.

Fritzell, J., Rehnberg, J., Hertzman, J. B., & Blomgren, J. (2015). Absolute or relative? A comparative analysis of the relationship between poverty and mortality. *International journal of public health*, *60*(1), 101-110.

Frohlich, K. L., Corin, E., and Potvin, L. (2001). A theoretical proposal for the relationship between context and disease. *Sociology of Health & Illness*, *23*(6), 776–797.

Gal, J. (2010). Is there an extended family of Mediterranean welfare states? *Journal of European Social Policy*, *20*(4), 283–300.

Gallie, D., and Paugam, S. (2000). *Welfare regimes and the experience of unemployment in Europe*. Oxford, UK; New York: Oxford University Press.

Geddes, B. (1990). How the cases you choose affect the answers you get: Selection bias in comparative politics. *Political Analysis*, *2*(1), 131–150.

Geyer, S., Hemström, Ö., Peter, R., and Vågerö, D. (2006). Education, income, and occupational class cannot be used interchangeably in social epidemiology. Empirical evidence against a common practice. *Journal of Epidemiology and Community Health*, *60*(9), 804–810.

Goldthorpe, J. H. (2010). Analysing Social Inequality: A Critique of Two Recent Contributions from Economics and Epidemiology. *European Sociological Review*, *26*(6), 731–744.

Goodin, R. E., Headey, B., Muffels, R., and Dirven, H.-J. (1999). *The Real Worlds of Welfare Capitalism*. Cambridge, UK: Cambridge University Press.

Goodman, A. H. (2000). Why genes don't count (for racial differences in health). *American Journal of Public Health*, *90*(11), 1699–1702.

Goodman, R., Kwon, H.-J. (1998). *The East Asian Welfare Model: Welfare Orientalism and the State*. London; New York: Routledge.

Gordon, R. A. (2012). *Applied Statistics for the Social and Health Sciences*. London; New York: Routledge.

Gottschalk, P., and Smeeding, T. M. (2000). Empirical evidence of income inequality in industrialized countries. In *The handbook of income distribution*. Amsterdam; New York: Elsevier.

Gough, D., Oliver, S., and Thomas, J. (2012). *An introduction to systematic reviews*. London; Thousand Oaks, Calif.: Sage.

Graham, H. (1995). Cigarette Smoking: A Light on Gender and Class Inequality in Britain? *Journal of Social Policy*, *24*(4), 509–527.

Gravelle, H. (1998). How much of the relation between population mortality and unequal distribution of income is a statistical artefact? *BMJ*, *316*(7128), 382–385.

Gravelle, H., Wildman, J., and Sutton, M. (2002). Income, income inequality and health: what can we learn from aggregate data? *Social Science & Medicine*, *54*(4), 577–589.

Gray, R. (2009, March 14). Nordic diet as healthy as Mediterranean foods. *The Daily Telegraph*. Retrieved from http://www.telegraph.co.uk/lifestyle/wellbeing/dietandfitness/4990538/Nordic-diet-as-healthy-as-Mediterranean-foods.html

Hair, J. F., Black, B., Babin, B., Anderson, R. E., and Tatham, R. L. (2005). *Multivariate Data Analysis* (6th edition). Upper Saddle River, NJ: Pearson.

Harding, S., Lenguerrand, E., Costa, G., d'Errico, A., Martikainen, P., Tarkiainen, L., … Bartley, M. (2013). Trends in mortality by labour market position around retirement ages in three European countries with different welfare regimes. *International Journal of Public Health*, *58*(1), 99–108.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, 1251–1271.

Hausmann, R. (2014). *Gender Gap Report 2014*. Cologny, Switzerland: World Economic Forum. Retrieved from http://reports.weforum.org/global-gender-gap-report-2014/

Helliwell, J. F., and Putnam, R. D. (2004). The social context of well-being. *Philosophical Transactions-Royal Society of London Series B Biological Sciences*, 1435–1446.

Herrnstein, R. and Murray, C. (1996) *Bell Curve: Intelligence and Class Structure in American Life.* New York: Free Press.

Hertz, E., Hebert, J. R., and Landon, J. (1994). Social and environmental factors and life expectancy, infant mortality, and maternal mortality rates: Results of a cross-national comparison. *Social Science & Medicine*, *39*(1), 105–114.

Hicks, A. (1994). Introduction to pooling. In Janoski, T. and Hicks, A. (Eds.) *The comparative political economy of the welfare state* (pp. 169–188). Cambridge, U.K.: Cambridge University Press.

Higgins, J. P. T., and Green, S. (2008). *Cochrane Handbook for Systematic Reviews of Interventions* (1 edition). Chichester, U.K.: Wiley.

Hoffmeister, H., Schelp, F.-P., Mensink, G. B., Dietz, E., and Böhning, D. (1999). The relationship between alcohol consumption, health indicators and mortality in the German population. *International Journal of Epidemiology*, *28*(6), 1066–1072.

Holliday, I. (2000). Productivist welfare capitalism: Social policy in East Asia. *Political studies*, *48*(4), 706-723.

Holliday, I. (2005). East Asian social policy in the wake of the financial crisis: farewell to productivism? *Policy & Politics*, *33*(1), 145–162.

Hood, A., and Johnson, P. (2014). What is welfare spending?, Research Paper. London, U.K.: Institute For Fiscal Studies. Retrieved from http://www.ifs.org.uk/publications/7424

Hothorn, T., Zeileis, A., Millo, G., and Mitchell, D. (2010). lmtest: Testing Linear Regression Models. *R Package Version 0.9-27*. Retrieved from http://cran.open-source-solution.org/web/packages/lmtest/

Huber, E., and Stephens, J. D. (2001). *Development and Crisis of the Welfare State: Parties and Policies in Global Markets*. Chicago: University of Chicago Press.

Hudson, J. (2012). Welfare Regimes and Global Cities: A Missing Link in the Comparative Analysis of Welfare States? *Journal of Social Policy*, *41*(3), 455–473.

Hudson, J., and Hwang, G.-J. (2013). Pathways of welfare state development in East Asia. In M. Izuhara (Ed.), *Handbook on East Asian Social Policy* (pp. 15-40). Cheltenham, UK: Edward Elgar Publishing.

Hudson, J., and Kühner, S. (2009). Towards productive welfare? A comparative analysis of 23 OECD countries. *Journal of European Social Policy*, *19*(1), 34–46.

Hug, S. (2003). Selection Bias in Comparative Research: The Case of Incomplete Data Sets. *Political Analysis*, *11*(3), 255–274.

Huijts, T., and Eikemo, T. A. (2009). Causality, social selectivity or artefacts? Why socioeconomic inequalities in health are not smallest in the Nordic countries. *The European Journal of Public Health*, *19*(5), 452–453.

Hurrelmann, K., Rathmann, K., and Richter, M. (2010). Health inequalities and welfare state regimes. A research note. *Journal of Public Health*, *19*(1), 3–13.

Ikeda, N., Saito, E., Kondo, N., Inoue, M., Ikeda, S., Satoh, T., … Shibuya, K. (2011). What has made the population of Japan healthy? *The Lancet*, *378*(9796), 1094–1105.

Illich, I. (1976). *Medical nemesis: the expropriation of health*. New York: Pantheon Books.

Janssen, F., Mackenbach, J. P., and Kunst, A. E. (2004). Trends in old-age mortality in seven European countries, 1950–1999. *Journal of Clinical Epidemiology*, *57*(2), 203–216.

Jen, M. H., Jones, K., and Johnston, R. (2009a). Compositional and contextual approaches to the study of health behaviour and outcomes: Using multi-level modelling to evaluate Wilkinson's income inequality hypothesis. *Health & Place*, *15*(1), 198–203.

Jen, M. H., Jones, K., and Johnston, R. (2009b). Global variations in health: evaluating Wilkinson's income inequality hypothesis using the World Values Survey. *Social Science & Medicine*, *68*(4), 643–653.

Jenkins, S. P. (2015). World Income Inequality Databases: an assessment of WIID and SWIID. *The Journal of Economic Inequality*, *13*(4), 629-671.

Jensen, C. (2008). Worlds of welfare services and transfers. *Journal of European Social Policy*, *18*(2), 151–162.

Jones, C. (1990). Hong Kong, Singapore, South Korea, and Taiwan: Oikonomic welfare states. *Government and Opposition* 25(3): 446–462.

Joumard, I., Pisu, M., and Bloch, D. (2012). Tackling income inequality: The role of taxes and transfers. *OECD Journal: Economic Studies*, *2012*(1), 37–70.

Judge, K., Mulligan, J.-A., and Benzeval, M. (1998). Income inequality and population health. *Social Science & Medicine*, *46*(4–5), 567–579.

Kääriäinen, J., and Lehtonen, H. (2006). The Variety of social capital in welfare state regimes - a comparative study of 21 countries. *European Societies*, *8*(1), 27–57.

Kangas, O. (2010). One hundred years of money, welfare and death: mortality, economic growth and the development of the welfare state in 17 OECD countries 1900-2000. *International Journal of Social Welfare*, *19*, S42–S59.

Karim, S. A., Eikemo, T. A., and Bambra, C. (2010). Welfare state regimes and population health: Integrating the East Asian welfare states. *Health Policy*, *94*(1), 45–53.

Kashin, K. (2014). Package panelAR; Retrieved from http://cran-mirror.cs.uu.nl/web/packages/panelAR/panelAR.pdf

Kavanagh, A. M., Turrell, G., and Subramanian, S. V. (2006). Does area-based social capital matter for the health of Australians? A multilevel analysis of self-rated health in Tasmania. *International Journal of Epidemiology*, *35*(3), 607–613.

Kawachi, I., and Kennedy, B. P. (1999). Income inequality and health: pathways and mechanisms. *Health Services Research*, *34*(1 Pt 2), 215.

Kawachi, I., Kennedy, B. P., Lochner, K., and Prothrow-Stith, D. (1997). Social capital, income inequality, and mortality. *American Journal of Public Health*, *87*(9), 1491–1498.

Kawachi, I., Subramanian, S. V., and Almeida-Filho, N. (2002). A glossary for health inequalities. *Journal of Epidemiology and Community Health*, *56*(9), 647–652.

Kennedy, B. P., Kawachi, I., Glass, R., and Prothrow-Stith, D. (1998). Income distribution, socioeconomic status, and self rated health in the United States: multilevel analysis. *BMJ*, *317*(7163), 917–921.

Kennedy, B. P., Kawachi, I., Prothrow-Stith, D., and others. (1996). Income distribution and mortality: cross sectional ecological study of the Robin Hood index in the United States. *BMJ*, *312*(7037), 1004–1007.

Kennelly, B., O'Shea, E., and Garvey, E. (2003). Social capital, life expectancy and mortality: a cross-national examination. *Social Science & Medicine*, *56*(12), 2367–2377.

Kersbergen, K. van, and Vis, B. (2015). Three worlds' typology: Moving beyond normal science? *Journal of European Social Policy*, *25*(1), 111–123.

Khang, Y. H., Cho, S. I., Yang, S., and Lee, M. S. (2005). Socioeconomic differentials in health and health related behaviors: findings from the Korea Youth Panel Survey. *Journal of Preventive Medicine and Public Health= Yebang Uihakhoe Chi*, *38*(4), 391–400.

Ki, M. (2009). *Health selection and health inequalities*. Doctoral Thesis. UCL (University College London). Retrieved from http://eprints.ucl.ac.uk/18913/

Kildal, N., and Kuhnle, S. (2007). *Normative Foundations of the Welfare State: The Nordic Experience*. London; New York: Routledge.

Kim, D., Baum, C. F., Ganz, M. L., Subramanian, S. V., and Kawachi, I. (2011). The contextual effects of social capital on health: A cross-national instrumental variable analysis. *Social Science & Medicine*, *73*(12), 1689–1697.

Kim, K.-t. (2015). From Worlds to Cases: Case Selection and “Other Worlds” in the Welfare Modelling Business. *Social Policy and Society*, *14*(2), 309–321.

Kim, P. H. (2010). The East Asian welfare state debate and surrogate social policy: an exploratory study on Japan and South Korea. *Socio-Economic Review*, *8*(3), 411–435.

Kim, Y.-M. (2011) Still Welfare Exceptionalism in East Asia?, (in Korean), *The Journal of Asiatic Studies*, 54(1), 7-22.

Kindig, D. A. (2007). Understanding population health terminology. *Milbank Quarterly*, *85*(1), 139–161.

Kindig, D., and Stoddart, G. (2003). What Is Population Health? *American Journal of Public Health*, *93*(3), 380–383.

Kittel, B. (1999). Sense and sensitivity in pooled analysis of political data. *European Journal of Political Research*, *35*(2), 225–253.

Kleiber, C., and Zeileis, A. (2008). *Applied Econometrics with R* (2008 edition). New York: Springer.

Kmenta, J. (1971). *Elements of econometrics*. Macmillan: New York.

Kondo, N., Sembajwe, G., Kawachi, I., van Dam, R. M., Subramanian, S. V., and Yamagata, Z. (2009). Income inequality, mortality, and self rated health: meta-analysis of multilevel studies. *BMJ : British Medical Journal*, *339*.

Korpi, W., and Palme, J. (1998). The Paradox of Redistribution and Strategies of Equality: Welfare State Institutions, Inequality, and Poverty in the Western Countries. *American Sociological Review*, *63*(5), 661–687.

Kunst, A. E., Bos, V., Lahelma, E., Bartley, M., Lissau, I., Regidor, E., … Mackenbach, J. P. (2005). Trends in socioeconomic inequalities in self-assessed health in 10 European countries. *International Journal of Epidemiology*, *34*(2), 295–305.

Kvist, J., Fritzell, J., Hvinden, B., and Kangas, O. (2012). Nordic Response to rising inequalities: still pursuing a distinct path or joining the rest? In J. Kvist, J. Fritzell, B. Hvinden, O. Kangas, and B. Hvinden (Eds.), *Changing Social Equality: The Nordic Welfare Model in the 21st Century* (pp. 201–206). Bristol, UK: Policy Press.

Kvist, J., & Greve, B. (2011). Has the Nordic welfare model been transformed?. *Social Policy & Administration*, *45*(2), 146-160.

Kwon, H.-J. (1997). Beyond European Welfare Regimes: Comparative Perspectives on East Asian Welfare Systems. *Journal of Social Policy*, *26*(4), 467–484.

Kwon, H. J. (2009). The reform of the developmental welfare state in East Asia. *International Journal of Social Welfare*, *18*(s1), S12-S21.

Lahelma, E., and Arber, S. (1994). Health inequalities among men and women in contrasting welfare states. *The European Journal of Public Health*, *4*(3), 213–226.

Lahelma, E., and Lundberg, O. (2009). Health inequalities in European welfare states. *The European Journal of Public Health*, *19*(5), 445–446.

Lahelma, E., Martikainen, P., Laaksonen, M., and Aittomäki, A. (2004). Pathways between socioeconomic determinants of health. *Journal of Epidemiology and Community Health*, *58*(4), 327–332.

Lawrence, D., Mitrou, F., and Zubrick, S. R. (2009). Smoking and mental illness: results from population surveys in Australia and the United States. *BMC Public Health*, *9*(1), 285.

Lee, M. R., and Bankston, W. B. (1999). political structure, economic inequality,and homicide: a cross-national analysis. *Deviant Behavior*, *20*(1), 27–55.

Lee, S. S. (2011). The Evolution of Welfare Production Regimes in East Asia: A Comparative Study of Korea, Japan, and Taiwan. *The Korean Journal of Policy Studies*, *26*(1), 49–75.

Lee, Y.-J., and Ku, Y. (2007). East Asian Welfare Regimes: Testing the Hypothesis of the Developmental Welfare State. *Social Policy & Administration*, *41*(2), 197–212.

Lester, D. (1987). Relation of income inequality to suicide and homicide rates. *Journal of Social Psychology*, *127*(1), 101–102.

Lindbom, A. (2001). Dismantling the Social Democratic welfare model? Has the Swedish welfare state lost its defining characteristics? *Scandinavian Political Studies*, *24*(3), 171–193.

Lindman, H. R. (2011). *Analysis of Variance in Experimental Design* (Softcover reprint of the original 1st ed. 1992 edition). New York: Springer.

Lindstrom, C., and Lindstrom, M. (2006). "Social capital," GNP per capita, relative income, and health: An ecological study of 23 countries. *International Journal of Health Services*, *36*(4), 679–696.

Lipscomb, H. J., Loomis, D., McDonald, M. A., Argue, R. A., and Wing, S. (2006). A Conceptual Model of Work and Health Disparities in the United States. *International Journal of Health Services*, *36*(1), 25–50.

Löckenhoff, C. E., Lee, D. S., Buckner, K. M., Moreira, R. O., Martinez, S. J., and Sun, M. Q. (2015). Cross-Cultural Differences in Attitudes About Aging: Moving Beyond the East-West Dichotomy. In Cheng, S., Chi, I., Fung, H., Li, L., Woo, J. (Eds.) *Successful Aging* (pp. 321–337). London: Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-94-017-9331-5_19

Luebker, M. (2012). The impact of taxes and transfers on inequality. Travail Policy Brief No. 4. Geneva: International Labour Office.

Lukaschek, K., Erazo, N., Baumert, J., and Ladwig, K.-H. (2012). Suicide Mortality in Comparison to Traffic Accidents and Homicides as Causes of Unnatural Death. An Analysis of 14,441 Cases in Germany in the Year 2010. *International Journal of Environmental Research and Public Health*, *9*(12), 924–931.

Lundberg, O. (1991). Childhood Living Conditions, Health Status, and Social Mobility: A Contribution to the Health Selection Debate. *European Sociological Review*, *7*(2), 149–162.

Lundberg, O., Yngwe, M. Å., Stjärne, M. K., Elstad, J. I., Ferrarini, T., Kangas, O., … Fritzell, J. (2008). The role of welfare state principles and generosity in social policy programmes for public health: an international comparative study. *The Lancet*, *372*(9650), 1633–1640.

Lupton, R. (2003). Neighbourhood effects: can we measure them and does it matter? *LSE STICERD Research Paper No. CASE073*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1158964

Lynch, J., Smith, G.D., Harper, S., Hillemeier, M., Ross, N., Kaplan, G. A., and Wolfson, M. (2004). Is Income Inequality a Determinant of Population Health? Part 1. A Systematic Review. *Milbank Quarterly*, *82*(1), 5–99.

Lynch, J., Smith, G.D., Hillemeier, M., Shaw, M., Raghunathan, T., and Kaplan, G. (2001). Income inequality, the psychosocial environment, and health: comparisons of wealthy nations. *The Lancet*, *358*(9277), 194–200.

Lynch, J. W., Smith, G.D., Kaplan, G. A., and House, J. S. (2000). Income inequality and mortality: importance to health of individual income, psychosocial environment, or material conditions. *BMJ*, *320*(7243), 1200–1204.

Lynch, J. W., and Kaplan, G. A. (1997). Understanding How Inequality in the Distribution of Income Affects Health. *Journal of Health Psychology*, *2*(3), 297–314.

Lynch, J. W., Kaplan, G. A., Pamuk, E. R., Cohen, R. D., Heck, K. E., Balfour, J. L., and Yen, I. H. (1998). Income inequality and mortality in metropolitan areas of the United States. *American Journal of Public Health*, *88*(7), 1074–1080.

Macinko, J. A., Shi, L., Starfield, B., and Wulu, J. T. (2003). Income inequality and health: a critical review of the literature. *Medical Care Research and Review*, *60*(4), 407–452.

Macinko, J. A., Shi, L. Y., and Starfield, B. (2004). Wage inequality, the health system, and infant mortality in wealthy industrialized countries, 1970-1996. *Social Science and Medicine*, *58*(2), 279–292.

Macintyre, S. (1997). The Black Rreport and beyond what are the issues? *Social Science & Medicine*, *44*(6), 723–745.

Macintyre, S., Ellaway, A., and Cummins, S. (2002). Place effects on health: how can we conceptualise, operationalise and measure them? *Social Science & Medicine*, *55*(1), 125–139.

Mackenbach, J. P. (2003). An analysis of the role of health care in reducing socioeconomic inequalities in health: The case of the Netherlands, *International Journal of Health Services*, *33*(3), 523~541.

Mackenbach, J. P. (2011). Public health and welfare. *The European Journal of Public Health*, *22*(1), 1–1.

Mackenbach, J. P. (2012). The persistence of health inequalities in modern welfare states: The explanation of a paradox. *Social Science & Medicine*, *75*(4), 761–769.

Mackenbach, J. P., Kunst, A. E., Cavelaars, A. E., Groenhof, F., and Geurts, J. J. (1997). Socioeconomic inequalities in morbidity and mortality in western Europe. *The Lancet*, *349*(9066), 1655–1659.

Mackenbach, J. P., Stirbu, I., Roskam, A.-J. R., Schaap, M. M., Menvielle, G., Leinsalu, M., and Kunst, A. E. (2008). Socioeconomic Inequalities in Health in 22 European Countries. *New England Journal of Medicine*, *358*(23), 2468–2481.

Mansyur, C., Amick, B. C., Harrist, R. B., and Franzini, L. (2008). Social capital, income inequality, and self-rated health in 45 countries. *Social Science & Medicine*, *66*(1), 43–56.

Marmot, M., and Smith, G. D. (1989). Why are the Japanese living longer? *BMJ*, *299*(6715), 1547–1551.

Marmot, M. G., Allen, J., Goldblatt, P., Boyce, T., McNeish, D., Grady, M., and Geddes, I. (2010). *Fair society, healthy lives: Strategic review of health inequalities in England post-2010.* London: Marmot Review Retrieved from http://eprints.ucl.ac.uk/111743/

Marmot, M., Ryff, C. D., Bumpass, L. L., Shipley, M., and Marks, N. F. (1997). Social inequalities in health: next questions and converging evidence. *Social Science and Medicine*, *44*(6), 901-910.

Marmot, M., and Wilkinson, R. G. (2001). Psychosocial and material pathways in the relation between income and health: a response to Lynch et al. *BMJ*, *322*(7296), 1233–1236.

McKeown, T. (1979). *The Role of Medicine: Dream, Mirage, Or Nemesis?* Princeton, NJ: Princeton University Press.

Mellor, J. M., and Milyo, J. (2001). Reexamining the evidence of an ecological association between income inequality and health. *Journal of Health Politics, Policy and Law*, *26*(3), 487–522.

Menon-Johansson, A. (2005). Good governance and good health: The role of societal structures in the human immunodeficiency virus pandemic. *BMC International Health and Human Rights*, *5*, 4.

Mills, C. (2012). Open letter to Prof. Richard Wilkinson and Prof. Kate Pickett [blogspot]. Retrieved from http://oxfordsociology.blogspot.co.uk/2012/11/open-letter-to-prof-richard-wilkinson.html

Muller, A. (2002). Education, income inequality, and mortality: a multiple regression analysis. *BMJ : British Medical Journal*, *324*(7328), 23.

Muller, E. N. (1985). Income inequality, regime repressiveness, and political violence. *American Sociological Review*, *50*(1), 47–61.

Muntaner, C., Borrell, C., Ng, E., Chung, H., Espelt, A., Rodriguez-Sanz, M., … O'Campo, P. (2011). Review article: Politics, welfare regimes, and population health: controversies and evidence. *Sociology of Health & Illness*, *33*(6), 946–964.

Muntaner, C., Lynch, J. W., Hillemeier, M., Lee, J. H., David, R., Benach, J., and Borrell, C. (2002). Economic inequality, working-class power, social capital, and cause-specific mortality in wealthy countries. *International Journal of Health Services*, *32*(4), 629–656.

Navarro, V., and Shi, L. (2001). The political context of social inequalities and health. *International Journal of Health Services*, *31*(1), 1–21.

Nazroo, J. Y. (1997). *The health of Britain's ethnic minorities: findings from a national survey.* London: Policy Studies Institute. Retrieved from http://discovery.ucl.ac.uk/179498/

Nelson, K., & Fritzell, J. (2014). Welfare states and population health: The role of minimum income benefits for mortality. *Social Science & Medicine*, *112*, 63-71.

Niedzwiedz, C. L., Katikireddi, S. V., Pell, J. P., and Mitchell, R. (2014). The association between life course socioeconomic position and life satisfaction in different welfare states: European comparative study of individuals in early old age. *Age and Ageing*, 43(3), 431-436.

Nolte, E., and McKee, M. (2004). *Does health care save lives?: avoidable mortality revisited*. London: Nuffield Trust.

Nordenmark, M., Strandh, M., and Layte, R. (2006). The impact of unemployment benefit system on the mental well-being of the unemployed in Sweden, Ireland and Great Britain. *European Societies*, *8*(1), 83–110.

Nowatzki, N. R. (2012). Wealth Inequality and Health: A Political Economy Perspective. *International Journal of Health Services*, *42*(3), 403–424.

Obinger, H., and Wagschal, U. (2001). Families of nations and public policy. *West European Politics*, *24*(1), 99–114.

OECD. (2008). *OECD Regions at a Glance 2007*. Paris: OECD.

OECD (2011). *Divided we stand: why inequality keeps rising*. Paris: OECD.

OECD (2012) *Counrtry Note: Sweden*, *Education at a Glance: OECD Indicators 2012,* Paris, OECD.

OECD. (2013a). *Health at a Glance 2013 OECD Indicators.* Paris: OECD. Retrieved from https://login.libproxy.uregina.ca:8443/login?url=http://dx.doi.org/10.1787/health_glance-2013-en

OECD (2013b). *OECD Factbook 2013: Economic, Environmental and Social Statistics*, Paris: OECD Publishing.

OECD (2013c), "Smoking", in *OECD Factbook 2013: Economic,Environmental and Social Statistics*. Paris: OECD Publishing. http://dx.doi.org/10.1787/factbook-2013-98-en

OECD (2014a) Society at a glance 2014 OECD social indicators, Paris: OECD. Retrieved from http://www.oecd.org/social/societyataglance.htm

OECD (2014b). *OECD Factbook 2014: Economic, Environmental, and Social Statistics*. Paris: OECD.

OECD (2014c). OECD Health Statistics 2014 Definitions, Sources and Methods. Retrieved from http://stats.oecd.org/wbos/fileview2.aspx?IDFile=9dc6e690-fa1d-40fa-a279-b726e2616433

OECD (2014d). *OECD Regions at a Glance 2013*. Paris: OECD.

OECD (2015a). OECD Better Life Index. Retrieved from http://www.oecdbetterlifeindex.org/topics/environment/

OECD (2015b) OECD.Stat homepage. Retrieved from http://stats.oecd.org/Index.aspx?DataSetCode=PDB_LV

OECD (2015c) OECD iLibrary. Retrieved from http://www.oecd-ilibrary.org/social-issues-migration-health/alcohol-consumption_alcoholcons-table-en

OECD (2015d) OECD Regional Demographic Statistics. Retrieved from https://stats.oecd.org/Index.aspx?DataSetCode=REG_DEMO_TL2#

OECD (2015e) *How's Life? 2015: Measuring Well-being.* Paris: OECD Publishing.

OECD (2015f) *Health at a Glance 2015: OECD Indicators*. Paris: OECD Publishing.

OECD (2016a). Labour Force Statistics. Retrieved from http://stats.oecd.org/Index.aspx?DatasetCode=LFS_SEXAGE_I_R#

OECD (2016b) OECD Data Homepage, retrieved from https://data.oecd.org/pop/population.htm

Oeppen, J., and Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, *296*(5570), 1029–1031.

Office for National Statistics. (2014). Death Registration Summary Tables, England and Wales, 2013. Retrieved from http://www.ons.gov.uk/ons/rel/vsob1/death-reg-sum-tables/2013/index.html

Olafsdottir, S. (2007). Fundamental causes of health disparities: stratification, the welfare state, and health in the United States and Iceland. *Journal of Health and Social Behavior*, *48*(3), 239–253.

Olshansky, S. J., Carnes, B. A., and Désesquelles, A. (2001). Prospects for Human Longevity. *Science*, *291*(5508), 1491–1492.

O'Neill, M. S., Jerrett, M., Kawachi, I., Levy, J. I., Cohen, A. J., Gouveia, N., … Schwartz, J. (2003). Health, wealth, and air pollution: advancing theory and methods. *Environmental Health Perspectives*, *111*(16), 1861–1870.

Osborne, J. W., and Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research & Evaluation*, *9*(6), 1–12.

Osler, M., Prescott, E., Christensen, U., Due, P., Engholm, G., and others. (2002). Income inequality, individual income, and mortality in Danish adults: analysis of pooled data from two cohort studies. *BMJ*, *324*(7328), 13.

Pampel, F. C. (2002). Inequality, diffusion, and the status gradient in smoking. *Social Problems*, *49*(1), 35–57.

Pampel, F. C., and Rogers, R. G. (2004). Socioeconomic status, smoking, and health: a test of competing theories of cumulative advantage. *Journal of Health and Social Behavior*, *45*(3), 306–321.

Park, C. U., & Jung, D. (2009). Making sense of the Asian welfare regimes with the western typology. *Korean Journal of Sociology*, *43*(3), 57-85.

Parks, R. W. (1967). Efficient Estimation of a System of Regression Equations when Disturbances are Both Serially and Contemporaneously Correlated. *Journal of the American Statistical Association*, *62*(318), 500–509.

Pearce, N., Foliaki, S., Sporle, A., and Cunningham, C. (2004). Genetics, race, ethnicity, and health. *BMJ*, *328*(7447), 1070–1072.

Peng, I., & Wong, J. (2008). Institutions and institutional purpose: continuity and change in East Asian social policy. *Politics & Society*, *36*(1), 61-88.

Pes, G. M., Lio, D., Carru, C., Deiana, L., Baggio, G., Franceschi, C., … Caruso, C. (2004). Association between longevity and cytokine gene polymorphisms. A study in Sardinian centenarians. *Aging Clinical and Experimental Research*, *16*(3), 244–248.

Pesaran, M. H. (2004). *General Diagnostic Tests for Cross Section Dependence in Panels* (SSRN Scholarly Paper No. ID 572504). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=572504

Petticrew, M., and Roberts, H. (2008). *Systematic reviews in the social sciences: A practical guide*. Chichester, U.K.: Wiley-Blackwell.

Piacentini, M. (2014), "Measuring Income Inequality and Poverty at the Regional Level in OECD Countries", *OECD Statistics Working Papers*, 2014/03, OECD Publishing.

Pickett, K. E., Kelly, S., Brunner, E., Lobstein, T., and Wilkinson, R. G. (2005). Wider income gaps, wider waistbands? An ecological study of obesity and income inequality. *Journal of Epidemiology and Community Health*, *59*(8), 670–674.

Pickett, K. E., and Pearl, M. (2001). Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. *Journal of Epidemiology and Community Health*, *55*(2), 111–122.

Pickett, K. E., and Wilkinson, R. G. (2015). Income inequality and health: A causal review. *Social Science & Medicine*, *128*, 316–326.

Piketty, T. (2014). *Capital in the Twenty-First Century*. (A. Goldhammer, Trans.) (First Edition). Cambridge, MA: Belknap Press.

Ploubidis, G. B., Dale, C., and Grundy, E. (2012). Later life health in Europe: how important are country level influences? *European Journal of Ageing*, *9*(1), 5–13.

Plümper, T., Troeger, V. E., and Manow, P. (2005). Panel data analysis in comparative politics: Linking method to theory. *European Journal of Political Research*, *44*(2), 327–354.

Podesta, F. (2002). Recent Development of Quantitative Comparative Methodology: the Case of Pooled Time Series Cross-Section Analysis. DSS Papers SOC 3-02.

Poortinga, W. (2006). Social capital: An individual or collective resource for health? *Social Science & Medicine*, *62*(2), 292–302.

Pop, I. A., Ingen, E. van, and Oorschot, W. van. (2013). Inequality, Wealth and Health: Is Decreasing Income Inequality the Key to Create Healthier Societies? *Social Indicators Research*, *113*(3), 1025–1043.

Popham, F., Dibben, C., and Bambra, C. (2013). Are health inequalities really not the smallest in the Nordic welfare states? A comparison of mortality inequality in 37 countries. *Journal of Epidemiology and Community Health*, 67, 412-418.

Powell, M. (2004). Social Democracy in Europe: renewal or retreat? In G. Bonoli and M. Powell (Eds.), *Social Democratic Party Policies in Contemporary Europe*. London; New York: Routledge.

Powell, M. (2015). "A Re-Specification of the Welfare State": Conceptual Issues in "The Three Worlds of Welfare Capitalism." *Social Policy and Society*, 14(2), 247-258.

Powell, M., and Barrientos, A. (2004). Welfare regimes and the welfare mix. *European Journal of Political Research*, *43*(1), 83–105.

Powell, M., and Barrientos, A. (2015). Introduction: Twenty Five Years of the Welfare Modelling Business. *Social Policy and Society*, *14*(2), 241–245.

Powell, M., and Kim, K.-t.. (2014). The "Chameleon" Korean Welfare Regime. *Social Policy & Administration*, *48*(6), 626–646.

Prättälä, R., Hakala, S., Roskam, A.-J. R., Roos, E., Helmert, U., Klumbiene, J., … Kunst, A. E. (2009). Association between educational level and vegetable use in nine European countries. *Public Health Nutrition*, *12*(11), 2174–2182.

Preston, S. H. (1975). The Changing Relation between Mortality and level of Economic Development. *Population Studies*, *29*(2), 231–248.

Preston, S. H., Glei, D. A., and Wilmoth, J. R. (2010). A new method for estimating smoking-attributable mortality in high-income countries. *International Journal of Epidemiology*, *39*(2), 430–438.

Putnam, R. D., Leonardi, R., and Nanetti, R. Y. (1994). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.

Ram, R. (2006). Further examination of the cross-country association between income inequality and population health. *Social Science & Medicine*, *62*(3), 779–791.

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 350–371.

Raphael, D. (2013). The political economy of health promotion: part 2, national provision of the prerequisites of health. *Health Promotion International*, *28*(1), 112–132.

Raphael, D. (2014). Challenges to promoting health in the modern welfare state: The case of the Nordic nations. *Scandinavian journal of public health*, *42*(1), 7-17.

Raphael, D., and Bryant, T. (2006). Maintaining population health in a period of welfare state decline: political economy as the missing dimension in health promotion theory and practice. *Promotion and Education*, *13*(4), 236.

Rau, R., Soroko, E., Jasilionis, D., and Vaupel, J. W. (2008). Continued reductions in mortality at advanced ages. *Population and Development Review*, *34*(4), 747–768.

Ravenholt, R. T. (1990). Tobacco's Global Death March. *Population and Development Review*, *16*(2), 213–240.

Regidor, E. (2006). Social determinants of health: a veil that hides socioeconomic position and its relation with health. *Journal of Epidemiology and Community Health*, *60*(10), 896–901.

Regidor, E., Pascual, C., Martinez, D., Calle, M. E., Ortega, P., and Astasio, P. (2011). The role of political and welfare state characteristics in infant mortality: a comparative study in wealthy countries since the late 19th century. *International Journal of Epidemiology*, *40*(5), 1187–1195.

Reich, R. B. (2011). *Aftershock: The Next Economy and America's Future*. New York: Vintage Books.

Richter, M., Rathman, K., Gabhainn, S. N., Zambon, A., Boyce, W., and Hurrelmann, K. (2012). Welfare state regimes, health and health inequalities in adolescence: a multilevel study in 32 countries. *Sociology of Health & Illness*, *34*(6), 858–879.

Roberts, E. M. (1997). Neighborhood social environments and the distribution of low birthweight in Chicago. *American Journal of Public Health*, *87*(4), 597–603.

Rodgers, G. B. (1979). Income and inequality as determinants of mortality: An international cross-section analysis. *Population Studies*, *33*(2), 343–351.

Roos, G., Johansson, L., Kasmel, A., Klumbiené, J., and Prättälä, R. (2001). Disparities in vegetable and fruit consumption: European cases from the north to the south. *Public Health Nutrition*, *4*(1), 35–43.

Rose, R. (2000). How much does social capital add to individual health? *Social Science & Medicine*, *51*(9), 1421–1435.

Rosen, G. (1993). *A History of Public Health* (Expanded edition). Baltimore: Johns Hopkins University Press.

Ross, C. E. (2000). Walking, exercising, and smoking: does neighborhood matter? *Social Science & Medicine*, *51*(2), 265–274.

Ross, N. A., Dorling, D., Dunn, J. R., Henriksson, G., Glover, J., Lynch, J., and Weitoft, G. R. (2005). Metropolitan-income inequality and working-age mortality: A cross-sectional analysis using comparable data from five countries. *Journal of Urban Health-Bulletin of the New York Academy of Medicine*, *82*(1), 101–110.

Rostila, M. (2007). Social capital and health in European welfare regimes: a multilevel approach. *Journal of European Social Policy*, *17*(3), 223–239.

Rowlingson, K. (2011). *Does income inequality cause health and social problems?* York, U.K.: Joseph Rowntree Foundation.

Rowlingson, K., and MacKay, S. (2012). *Wealth and the wealthy: exploring and tackling inequalities between rich and poor*. Bristol, U.K.: Policy Press.

Rudat, K. (1994). *Black and minority ethnic groups in England: health and lifestyles. London: Health Education Authority*.

Runciman, W. (1966). *Relative deprivation and social justice: study attitudes social inequality in 20th century England*. London: Routledge Kegan Paul.

Russell, H. (2016, 2nd February). For pensioners, something is spot-on in the state of Denmark. *The Guardian*. Retrieved from http://www.theguardian.com/world/2016/feb/02/for-pensioners-something-is-spot-on-in-the-state-of-denmark

Sachs, B. P., Fretts, R. C., Gardner, R., Hellerstein, S., Wampler, N. S., and Wise, P. H. (1995). The impact of extreme prematurity and congenital anomalies on the interpretation of international comparisons of infant mortality. *Obstetrics & Gynecology*, *85*(6), 941–946.

Sacker, A., Worts, D., and McDonough, P. (2009). Social influences on trajectories of self-rated health: evidence from Britain, Germany, Denmark and the USA. *Journal of Epidemiology & Community Health*, *65*(2), 130–136.

Saint-Arnaud, S., and Bernard, P. (2003). Convergence or resilience? A hierarchical cluster analysis of the welfare regimes in advanced countries. *Current Sociology*, *51*(5), 499–527.

Sanders, A. E., Slade, G. D., John, M. T., Steele, J. G., Suominen-Taipale, A. L., Lahti, S., … Allen, P. F. (2009). A cross-national comparison of income gradients in oral health quality of life in four welfare states: application of the Korpi and Palme typology. *Journal of Epidemiology & Community Health*, *63*(7), 569–574.

Sanmartin, C. (2003). Labour market income inequality and mortality in North American metropolitan areas. *Journal of Epidemiology & Community Health*, *57*(10), 792–797.

Saunders, P. (2010). *Beware False Prophets: Equality, the Good Society and The Spirit Level.* London: Policy Exchange.

Sayers, A. (2007). Tips and tricks in performing a systematic review. *The British Journal of General Practice*, *57*(542), 759.

Scarborough, P., Bhatnagar, P., Wickramasinghe, K. K., Allender, S., Foster, C., and Rayner, M. (2011). The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006–07 NHS costs. *Journal of Public Health*, *33*(4), 527–535.

Schridde, H. (2002). Local Welfare Regimes and the Restructuring of the Welfare State an Anglo-German Comparison. *German Policy Studies/Politikfeldanalyse*, *2*(1), 105-142.

Schro¨der, M. (2008). 'Integrating welfare and production typologies: how refinements of the varieties of capitalism approach call for a combination of welfare typologies', *Journal of Social Policy*, 38(1),19–43.

Scruggs, A. and Allan, J. (2006) 'Welfare-state decommodification in 18 OECD countries: a replication and revision', *Journal of European Social Policy*, 16, 1, 55–72.

Scruggs, L. (2007). What's Multiple Regression got to do with it. *Comparative Social Research*, *24*, 309–323.

Sen, A. (2002). Health: perception versus observation: Self reported morbidity has severe limitations and can be extremely misleading. *BMJ*, *324*(7342), 860.

Seth, M. J. (2002). *Education Fever: Society, Politics, and the Pursuit of Schooling in South Korea*. Honolulu: University of Hawaii Press.

Shadish, W. (2006). Foreword. In Petticrew, M., and Roberts, H., *Systematic Review in the Social Sciences* (pp. vi–ix). Chichester, U.K.: Wiley-Blackwell.

Shalev, M. (1996). *The privatization of social policy? : occupational welfare and the welfare state in America, Scandinavia, and Japan*. Houndmills, Basingstoke, Hampshire; New York: Palgrave Macmillan.

Shalev, M. (2007). Limits and alternatives to multiple regression in comparative research. *Comparative Social Research*, *24*(3), 261–308.

Shapiro, S. S., and Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, *52*(3/4), 591–611.

Shaw, M., Dorling, D., Gordon, D., and Smith, G. D.(1999). *The widening gap: health inequalities and policy in Britain*. Bristol, UK: Policy Press.

Shi, L. (1992). The Relationship between Primary Care and Life Chances. *Journal of Health Care for the Poor and Underserved*, *3*(2), 321–335.

Shi, L., Macinko, J., Starfield, B., Politzer, R., and Xu, J. (2005). Primary care, race, and mortality in US states. *Social Science & Medicine*, *61*(1), 65–75.

Shibuya, K., Hashimoto, H., and Yano, E. (2002). Individual income, income distribution, and self rated health in Japan: cross sectional analysis of nationally representative sample. *BMJ*, *324*(7328), 16.

Shkolnikov, V. M., Andreev, E. M., Zhang, Z., Oeppen, J., and Vaupel, J. W. (2011). Losses of expected lifetime in the United States and other developed countries: Methods and empirical analyses. *Demography*, *48*(1), 211–239.

Siaroff, A. (1994). Work, welfare and gender equality: a new typology. In D. Sainsbury (Ed.), *Gendering Welfare States* (pp. 82–100). London: Sage Publications.

Simon, J. W., and Picazo, M. I. (2005). Reconfigurations of the Chilean State in Social Policy: the Emergence of a Post-neoliberal State. Presentation at International Sociology Association. Retrieved from http://works.bepress.com/jeanne_simon/2/

Sjöberg, O. (2014). Old-age pensions and population health: A global and cross-national perspective. *Global Public Health*, *9*(3), 271–285.

Smith, G. (Ed.). (2003). *Health Inequalities: Life Course Approaches*. Bristol, U.K.: Policy Press.

Smith, G., Blane, D., and Bartley, M. (1994). Explanations for socio-economic differentials in mortality. *European Journal of Public Health*, *4*(2), 131–144.

Smith, V., Devane, D., Begley, C. M., and Clarke, M. (2011). Methodology in conducting a systematic review of systematic reviews of healthcare interventions. *BMC Medical Research Methodology*, *11*(1), 15.

Solt, F. (2014). The Standardized World Income Inequality Database, SWIID Version 5.0. Retrieved from http://fsolt.org/swiid/

Solt, F. (2015). On the assessment and use of cross-national income inequality datasets. *The Journal of Economic Inequality*, *13*(4), 683-691.

Sorlie, P., Rogot, E., Anderson, R., Johnson, N. J., and Backlund, E. (1992). Black-white mortality differences by family income. *The Lancet*, *340*(8815), 346–350.

Staetsky, L. (2009). Diverging trends in female old-age mortality: A reappraisal - ProQuest. *Demographic Research*, *21*, 885–914.

Stanistreet, D., Bambra, C., and Scott-Samuel, A. (2005). Is patriarchy the source of men's higher mortality? *Journal of Epidemiology and Community Health*, *59*(10), 873–876.

Stansfeld, S. A., Clark, C., Rodgers, B., Caldwell, T., and Power, C. (2011). Repeated exposure to socioeconomic disadvantage and health selection as life course pathways to mid-life depressive and anxiety disorders. *Social Psychiatry and Psychiatric Epidemiology*, *46*(7), 549–558.

Starfield, B. (2007). Pathways of influence on equity in health. *Social Science & Medicine*, *64*(7), 1355–1362.

Starfield, B., Shi, L., and Macinko, J. (2005). Contribution of Primary Care to Health Systems and Health. *Milbank Quarterly*, *83*(3), 457–502.

Stier, H., Lewin-Epstein, N., and Braun, M. (2001). Welfare Regimes, Family-Supportive Policies, and Women's Employment along the Life-Course1. *American Journal of Sociology*, *106*(6), 1731–1760.

Stimson, J. A. (1985). Regression in Space and Time: A Statistical Essay. *American Journal of Political Science*, *29*(4), 914–947.

Subramanian, S. V., and Kawachi, I. (2003). Response: In defence of the income inequality hypothesis. *International Journal of Epidemiology*, *32*(6), 1037–1040.

Subramanian, S. V., and Kawachi, I. (2004). Income inequality and health: what have we learned so far? *Epidemiologic Reviews*, *26*(1), 78–91.

Subramanian, S. V., and Kawachi, I. (2006). Whose health is affected by income inequality? A multilevel interaction analysis of contemporaneous and lagged effects of state income inequality on individual self-rated health in the United States. *Health & Place*, *12*(2), 141–156.

Subramanian, S. V., Kawachi, I., and Kennedy, B. P. (2001). Does the state you live in make a difference? Multilevel analysis of self-rated health in the US. *Social Science & Medicine*, *53*(1), 9–19.

Sugiura, Y., Ju, Y.-S., Yasuoka, J., and Jimba, M. (2010). Rapid increase in Japanese life expectancy after World War II. *Biosci Trends*, *4*(1), 9–16.

Sung, K. (2001). Elder respect: exploration of ideals and forms in East Asia. *Journal of Aging Studies*, *15*(1), 13–26.

Takata, H., Ishii, T., Suzuki, M., Sekiguchi, S., and Iri, H. (1987). Influence of Major Histocompatibility Complex Region Genes on Human Longevity Among Okinawan-Japanese Centenarians and Nonagenarians.. *The Lancet*, *330*(8563), 824–826.

Takegawa, S. (2005), Japan’s welfare-state regime: welfare politics, provider and regulator, *Development and Society*, 34, 2: 169–90 (in Korean).

Tapia Granados, J. A. (2010). Politics and health in eight European countries: A comparative study of mortality decline under social democracies and right-wing governments. *Social Science & Medicine*, *71*(5), 841–850.

Tapia Granados, J. A. (2013). Health at advanced age: Social inequality and other factors potentially impacting longevity in nine high-income countries. *Maturitas*, *74*(2), 137–147.

Taxpayers Alliance. (2010). *The Spirit Illusion, A critical analysis of how “The Spirit Level” compares countries.* London: Taxpayers Alliance.

Theorell, T., and Vogel, J. (2003). Health in Vogel, J., (Ed.) *European welfare production: Institutional configuration and distributional outcome*. (pp. 147-170). New York: Springer.

Torre, R., and Myrskylä, M. (2014). Income inequality and population health: An analysis of panel data for 21 developed countries, 1975–2006. *Population Studies*, *68*(1), 1–13.

Torssander, J., and Erikson, R. (2010). Stratification and Mortality—A Comparison of Education, Class, Status, and Income. *European Sociological Review*, *26*(4), 465–474.

Townsend, P., and Davidson, N. (1982). The Black Report. In P. Townsend, N. Davidson, and M. Whitehead, *Inequalities in Health: the Black Report and the Health Divide* (pp. 31–216). London: Penguin.

Uphoff, E. P., Pickett, K. E., Cabieses, B., Small, N., and Wright, J. (2013). A systematic review of the relationships between social capital and socioeconomic inequalities in health: a contribution to understanding the psychosocial pathway of health inequalities. *Int J Equity Health*, *12*, 54.

Van De Mheen, H. D., Stronks, K., and Mackenbach, J. P. (1998). A Lifecourse Perspective on Socio-economic Inequalities in Health: The Influence of Childhood Socio-economic Conditions and Selection Processes. *Sociology of Health & Illness*, *20*(5), 754–777.

van der Gaag, M. P. J. (2005). *Measurement of Individual Social Capital*. Amsterdam: F&N Boekservices.

Vaupel, J. W., Zhang, Z., and Raalte, A. A. van. (2011). Life expectancy and disparity: an international comparison of life table data. *BMJ Open*, *1*(1).

Veenstra, G. (2002). Social capital and health (plus wealth, income inequality and regional health governance). *Social Science & Medicine*, *54*(6), 849–868.

Vogel, E. (1973) Japan as Number One: Lessons for America. Cambridge, MA: Harvard University Press.

Vrooman, J. C. (2012). Regimes and cultures of social security: Comparing institutional models through nonlinear PCA. *International Journal of Comparative Sociology*, *53*(5–6), 444–477.

Wagstaff, A., and van Doorslaer, E. (2000). Income Inequality and Health: What Does the Literature Tell Us? *Annual Review of Public Health*, *21*(1), 543–567.

Walker, A., and Wong, C-K. (2005). *East Asian welfare regimes in transition : from Confucianism to globalisation*. Bristol, UK: Policy Press.

Wawro, G. (2002). Estimating Dynamic Panel Data Models in Political Science. *Political Analysis*, *10*(1), 25–48.

Whelan, C. T., and Maitre, B. (2010). Welfare regime and social class variation in poverty and economic vulnerability in Europe: an analysis of EU-SILC. *Journal of European Social Policy*, *20*(4), 316–332.

Whitehead, M., and Dahlgren, G. (2001). Developing the Policy Response to Inequities in Health: A Global Perspective. In Evans, T., Whitehead, M., Diderichsen, F., Bhuiya, A., and Wirth, M. (Eds.), *Challenging Inequities in Health* (pp. 309–322). Oxford, U.K.: Oxford University Press.

Wilding, P. (2008). Is the East Asian Welfare Model still productive? *Journal of Asian Public Policy*, *1*(1), 18–31.

Wildman, J., Gravelle, H., and Sutton, M. (2003). Health and income inequality: attempting to avoid the aggregation problem. *Applied Economics*, *35*(9), 999–1004.

Wilkinson, R. G. (1992). Income distribution and life expectancy. *BMJ*, *304*(6820), 165–168.

Wilkinson, R. G. (1992). National mortality rates: the impact of inequality? *American Journal of Public Health*, *82*(8), 1082–1084.

Wilkinson, R. G. (1996). *Unhealthy societies : the afflictions of inequality*. London: New York: Routledge.

Wilkinson, R. G. (1999). The culture of inequality. In Kawachi, I., Kennedy, B., Wilkinson, R. G. (Eds.) *Income Inequality and Health.* (pp. 492–498). *The Society and Population Health Reader*. New York: The New Press.

Wilkinson, R. G., and Pickett, K. E. (2009). *The spirit level : why greater equality makes societies stronger*. New York: Bloomsbury Press.

Wilkinson, R. G., and Pickett, K. E. (2006). Income inequality and population health: a review and explanation of the evidence. *Social Science & Medicine (1982)*, *62*(7), 1768–1784.

Wilkinson, R. G., and Pickett, K. E. (2008). Income Inequality and Socioeconomic Gradients in Mortality. *American Journal of Public Health*, *98*(4), 699–704.

Wilkinson, R., and Marmot, M. (Eds.). (2003). *The solid facts: social determinants of health* (2nd ed). Copenhagen: WHO Regional Office for Europe.

Williams, G. H. (2003). The determinants of health: structure, context and agency. *Sociology of Health & Illness*, *25*(3), 131–154.

Wilmoth, J. R. (1998). Is the Pace of Japanese Mortality Decline Converging Toward International Trends? *Population and Development Review*, *24*(3), 593–600.

Wolfson, M., Gravelle, H., Wilkinson, R. G., Kaplan, G., Lynch, J., Ross, N., and Backlund, E. (1999). Relation between income inequality and mortality: empirical demonstration Diminishing returns to aggregate level studies Two pathways, but how much do they diverge? *BMJ*, *319*(7215), 953–957.

Wooldridge, J. M. (2001). *Econometric Analysis of Cross Section and Panel Data* (1 edition). Cambridge, MA: The MIT Press.

Wooldridge, J. M. (2008). *Introductory Econometrics: A Modern Approach* (4 edition). Mason, OH: Cengage Learning.

World Bank. (2010). *World Development Indicators 2010*. Washington D.C.: World Bank Publications.

World Bank. (2015). World Databank. Retrieved from http://data.worldbank.org/indicator/SL.UEM.TOTL.ZS

World Value Survey Association (2015) World Value Survey, retrieved from http://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp

Yunker, J. (2000) *Common Progress: The Case for a World Economic Equalization Program*, Westport, CT: Praeger Publisher.

Zheng, H. (2012). Do people die from income inequality of a decade ago? *Social Science & Medicine*, *75*(1), 36–45.