
Socioeconomic Inequality of Physical Activities in Canada

Mingying Zhu

(6748683)

Major Paper presented to the
Department of Economics of the University of Ottawa
in partial fulfillment of the requirements of the M.A. Degree

Supervisor: Professor Paul Makdissi

ECO 6099

December 2014

Abstract

In the field of health economics, the significance of physical activities to health is highlighted. Exercise is beneficial to improve health and can also help to prevent obesity. It is necessary to study the real exercise situation among population. The objective of this paper is to measure socioeconomic activity inequality among Canadians. The data are collected from the Canadian Community Health Survey 2012. Average daily energy expenditure serves as a good alternative to study the features of each physical activity. This paper mainly adopts a concentration index (*CI*), which indicates a socioeconomic dimension to exercise inequality, as well as an achievement index (*AI*) that includes mean value and presents a better inequality situation. This work also decomposes the *CI* and *AI* from two aspects: by categories of physical activities, and by demographic groups (gender, age, immigrant or not, the length of time since immigration). The results suggest that most of Canadians' exercise intensity is inactive or moderate. The poor prefer to choose some simple and single exercises like walking for exercise, gardening, bicycling and the like. The rich take more team sports, like ice hockey, tennis and baseball. The socioeconomic inequalities in exercise within subgroup show more significance.

Keywords: Socioeconomic Inequality; Physical Activity; Concentration Index;

Achievement Index

Table of Contents

1 Introduction	5
2 Literature Review	7
2.1 Literature on Physical Activity	7
2.2 Literature on Socioeconomic Health Inequality Measurement	10
3 Theoretical Framework	14
3.1 The Measurement of Physical Activities	14
3.2 Gini Coefficient and Lorenz Curve	16
3.3 Concentration Index and Concentration Curve	18
3.4 Achievement Index	23
3.5 Decomposition of Concentration Index	24
3.6 Decomposition of Achievement Index	27
4 Empirical Analysis	29
4.1 Data and Variables	29
4.2 Physical Activity Index	30
4.3 Decomposition by Categories of Physical Activities	32
4.4 Decomposition by Population Groups	39
5 Conclusion	44
6 Appendix	46
Table 1I: MET Table for Physical Activities	46
Table 1II: Total Daily Energy Expenditure	47

Table 2: Household Income Distribution from All Sources-----	47
Table 3: Physical Activity Distribution-----	48
Table 4: Concentration Index Decomposed by Various Physical Activities (Contribution of Each Activity) -----	48
Table 5: Concentration Index Decomposed by Demographic Groups -----	49
Table 6: Achievement Index Decomposed by Activities-----	50
Table 7: Achievement Index Decomposed by Demographic Groups-----	51
7 References -----	52

1. Introduction

As we all know, the appropriate amount of physical activities is very significant to a healthy body, as well as psychological wellbeing. Exercise could help participants consume overmuch calories so as to help lose weight or maintain a proper weight. It is also beneficial to improve heart and lung function, and also promote blood circulation. Therefore, the study of physical activity is significant to health economics. The purpose of our work is to analyze the Canadians' socioeconomic inequality in exercise.

The socioeconomic status is related to several factors, like education, income, family background, occupation and so on. We select household income within a year as a reasonable indicator to display socioeconomic position. At the same time, there are lots of sports related variables, like frequency and duration. A MET value, which indicates the energy expenditure of each activity, is a multiple of energy cost at the resting rate. We also include the MET to calculate the daily energy expenditure.

Wagstaff, Paci and van Doorslaer (1991) compare several kinds of inequality measurements, and state that a concentration index (*CI*) is one of the best ways to measure socioeconomic health inequality. Wagstaff (2002) also takes the mean value into consideration and then proposes the use of achievement index (*AI*). In our paper, we follow the recommendation from Wagstaff and select both *CI* and *AI*. Clarke, Gerdtham and Connelly (2003) analyze the *CI* in a more disaggregated way. They decompose the index from two different dimensions: one by various components; one by population groups according to various demographic characteristics. We closely

follow their method to decompose our *CI* by different types of physical activities and also by gender, age immigration status, and length of years since immigration respectively. Makdissi, Sylla and Yazbeck (2013) extend the decomposition application to achievement index. We also adopt their approach to disaggregate *AI* index from those two aspects above. With *CI* and *AI* index separately, we obtain the overall inequality situation among Canadians. After we conduct the decomposition, we could explain the largest contributor of a corresponding activity, as well as the effect of those demographic factors on socioeconomic inequality.

The remainder of the paper is organized as follows. The next section reviews the literature. This part is mainly divided into two types of a literature discussion: one for the studies related to physical activity, socioeconomic status and citizen health; another one for the influential works of the measurement of socioeconomic health inequality and its decomposition. The third section mainly introduces the related theoretical framework in detail. Section four presents the empirical results. The last section concludes the paper.

2. Literature Review

2.1 Literature on Physical Activity

According to Shield (2006), within 30 years from 1976 to 2006, the obesity and overweight among youth at the age from 12 to 17 have tripled in Canada. Moreover, around 75% of them will also stay obesity when they grow into the adults' age. The problem of obesity causes more public concern. It leads to a number of physiological disorders and diseases, such as cardiovascular disease, pulmonary function abnormalities, diabetes, musculoskeletal disease and endocrine disease.

Like our paper, Seliske, Pickett and Janssen (2012) also follow the obesity problem in terms of related active transportation and physical activities. They analyze the relationships between urban sprawl in Canada and the physical activity intensity. They find that there is a positive relationship between urban sprawl and the overall amount of exercise. However, there still exists a clear variation and inequality between youth group and adult group. Therefore, age can be taken into account in order to reduce the obesity in the process of urban sprawl. In our paper, we discuss the exercise both during the leisure time and transportation, and we put both two into the category of physical activities. We also discuss the inequality among different age groups. However, we focus more on the socioeconomic inequality of physical activities, and we decompose the physical activities by different sources and group while analyzing.

Another related work concerning socioeconomic inequality and physical activities is conducted by Beenackers et al. (2012). They measure a socioeconomic

status through comprehensive analysis of income, education and occupation. However, in our paper, we only select household income as an indicator of individual socioeconomic position. In their paper, when they discuss the inequality of physical activities, they pay more attention to the difference in time allocation and exercise intensity. Nevertheless, our work decomposes the physical activities into different sources, like swim, dance, jog, baseball, and so on, and we also study the inequality among different age groups, immigration groups and gender groups. They conclude that people at higher socioeconomic position tend to get more leisure time active physical activities, while people at lower socioeconomic position do more occupational physical activities.

There is lots of research related to physical activity and obesity. Dowler (2001) conducts a study concerning the inequalities in diet and physical activity in Europe. She find that the unemployed men and women are more sedentary. They could have more time to do the active physical exercise, but actually they get less physical activities and also have more proportion of obesity. In the case of the occupational group, she states that there is a lower inequality in men group for physical activities than women. Besides occupation and gender, education and age also matter. She also find that those with higher education background, they tend to engage in more active physical activities. Especially for age group, the result represents a positive relationship between education status and physical activities intensity and a negative relationship related to obesity.

Pan et al. (2009) conducted a cross-sectional research with respect to the

situation of physical activities among Canadians. The interesting part in their work is that when they study the influence factor of physical activities, they add “self-perceived benefit” (what you think can benefit from exercise) and “perceived barrier” (what makes you do not want to do the exercise). Meanwhile, they also take several traditional factors like income and education into consideration. In their results, they indicate that the male group seem to be more rigid in activities, while the old group is more sensitive to those four factors including the two perceived ones. Education has larger impact on the female group. The perceived barrier tends to have larger influence on the youth and the female. They also state that fundamental fitness facilities work better on those who get higher education.

For the times series data, odds ratio is a good choice to study the physical activity trends and between group inequality. Craig et al. (2004) write a report concerning twenty-year trends (from 1981 to 2000) in physical activity among Canadian adults. In their study, they use odds ratio to discuss and compare the different exercise trends of each group (age, education level and income status). Their results show that inequality gaps of those with different income and age become larger, while for those with different education position, exercise inequality gradually decreases. As for our paper, we abandon the method of odds ratio, as a result of our cross-sectional data and our different emphasis on inequality. Because the focus of our work is to analyze the socioeconomic inequality, we also pay attention to some influential works concerning the measurement of inequality.

2.2 Literature on Socioeconomic Health Inequality Measurement

Wagstaff et al. (1991) compare several representative methods of inequality measurement, such as the range, the Lorenz curve and Gini coefficient, pseudo Lorenz curves, the index of dissimilarity, the slope and relative indices of inequality and the concentration index. They point out that the method of “range” is too rough and radical. It ignores the possible change in the intermediate group and it also does not reflect the different group size. Wagstaff et al. (1991) state that the Gini coefficient does not capture the socioeconomic distribution of health inequality. Sampling data, selection size and sampling methods can easily affect the accuracy of index in the absence of all household income information. No matter the transfer of income occurs among the rich, among the poor, or between the rich and the poor, the Gini coefficient will be affected. Therefore, the Gini indices are sometimes misleading. The pseudo Lorenz curves and the index of dissimilarity cannot present the socioeconomic inequality of health status. They only reflect the health and related population share, while ignore the related socioeconomic components of each health part.

Wagstaff et al. (1991) recommend both the slope index and the concentration index (*CI*), which consider both health status and socioeconomic ranks. In our paper, we also adopt his method of concentration index as one of our main indicators. One of the related works using the *CI* is Zhang and Wang (2003) on socioeconomic inequality of obesity in the US. They give the clear and different definitions to obesity ($BMI \geq 30$) and overweight ($BMI \geq 25$), and also divide the agents according to various variables, like gender, age and ethnicity. They choose household income as their main

indicator of socioeconomic status, which is the same as what we select in our study. When they discuss the inequality, they conduct very careful decompositions. They decompose each gender group by age and ethnicity, and then they explain different *CI*s for male and female at different age group, as well as ethnicity group. The *CI* is almost negative for both gender groups, in which female's *CI* seems to be more serious. The *CI* gets smaller within the old over 50, while for those who are at 40-49, there is larger inequality among male group and female group. Moreover, there's no clear evidence showing that minorities suffer more socioeconomic inequality than the white.

Besides the concentration index, Wagstaff (2002) also introduces another useful measurement, the achievement index, in order to improve the role of the concentration index. *CI* can well explain the relationship between the socioeconomic position and ones' corresponding health status. Through the analysis of the *CI* curve radian and its distance from the diagonal, we can get one's degree of health inequality. The achievement index considers both inequality and mean value. After the deformation of *CI*, achievement index (*AI*) can be seen as a weighted average level, which automatically gives more weights to the poor but less weights to the rich. When Wagstaff (2002) study the health inequality, he also introduces the attitude to inequality into their formula, which reflects the aversion of one to health inequality. This idea is originally from Atkinson (1970), and then Yizhaki (1983) adopts the aversion variable (v) and extends the formula of the Gini index. The aversion variable (v) has been widely used in health inequality based on the deformation of Gini, *CI* and

AI. We also measure the socioeconomic inequality of physical activities on the basis of the concentration index and the achievement index.

Besides that, we also conduct our research based on the theory of decomposition. A separate study of *CI* sometimes seems to be too general, which needs decomposition to help analyze the relative contribution of its subgroup. Clarke et al. (2003) proposed to decompose the *CI* into two dimensions: one by components, and the other one by population characteristics. Firstly, they state that the average health status may include scores from both physical health and mental health, in which the different categories scores can be added together to get the general value. As for the whole *CI*, different parts may contribute differently to the overall *CI*. It is necessary to make a clear study and comparison of *CI* for its items. Like our study of physical activities, daily average energy expenditure is composed of daily average energy expenditure of various exercises, such as swimming, walk, dance, bicycling, tennis and so on. Through the discussion of each separate contribution to the overall *CI*, we can present the citizen's activities preference, as well as the relationship between socioeconomic status and each activity density. Secondly, Clarke et al. (2003) also put forward the other decomposition by population dimension, such as gender, age, ethnics and occupation. This kind of decomposition method is very powerful for the study of inequality. It explains both inequalities between group and within group, which helps to reflect the inequality problems of the targeted agents. Based on that, the policy strategies can propose more targeted suggestion suitable for specific population.

With respect to achievement index, Makdissi et al. (2013) put forward the corresponding method of decomposition by both components and population subgroups. In their study, they divide health problems into eleven types, including vision, hearing, pain, cognition, etc. They carry on the decomposition work of the achievement index for each health problem and then they obtain the relative contribution of corresponding health problem. With *CI* only, we can explain whether the specific health problem is equally distributed through the agents, or whether the poor suffer more from this disease. Besides that the decomposition index of achievement can explain not only the inequality situation but also the severity of each health attribute, which states clearly the comprehensive adverse effect of the targeted component (health attribute or activity sources).

In our paper, we calculate the overall achievement index for daily physical activities energy expenditure. Besides that, we also decompose the achievement index for each representative exercise and then work out the clear relative contribution. In regard to corresponding decomposition for various groups, Makdissi et al. (2013) introduce the population share for each group and also add a transvariation term into the formula. That transvariation term can be seen as a residual, according to what they explain in their work, which comes from the different socioeconomic positions of agent within his specific demographic group and within total population. They divide the population of the U.S. into Northeast, South, Midwest and West groups, and then they compare each health achievement of four subgroups. Their results suggest that when the aversion to inequality is high enough, the socioeconomic health inequality

between regions are the main contributors. In our paper, we use several traditional classification characteristics, like gender, age, immigration status.

3. Theoretical Framework

3.1 The Measurement of Physical Activities

Physical activity measurement should involve not only duration but also frequency. Meanwhile, different exercises may also show different fitness effects. For example, swimming and running for an hour will lead to different activity energy expenditure. Therefore, a simple index which represents daily, weekly or monthly exercise time cannot tell the whole story of the real situation of individual's sport intensity. We have to consider their relevant energy cost. Here we introduce MET (Metabolic Equivalent of Task) value to measure the energy expenditure of different exercise. According to Ainsworth et al. (1993), when they study the energy cost of people's exercise, they emphasize the definition of MET. MET value reflects metabolic energy expenditure, which is a multiple value of the energy consumption under the resting rate. For example, MET for swimming is 3. One who swims can spend three times energy as he is at rest. Because each sport could have various intensity under different exercise types, such as competition with vigorous effort, team games with moderate effort and leisure fun with light effort, we prefer the general value or stationary mean value. The MET values we use here are from CCHS (Canadian Community Health Survey). Table 1I in the appendix "MET Table for Physical Activities" lists the METs for 26 kinds of representative physical activities,

including walking for exercise, gardening or yard work, swimming, bicycling, popular or social dance etc. At the same time, besides the exercises during the leisure time, we also include the activity during the transportation, i.e. walking and bicycling to school or work.

In regard to physical activity measure, we take the average daily energy expenditure as our core analysis variable, which is required to calculate for each activity. Then adding each of the activity energy measures together gives the total daily average energy cost one spends. The relevant formula is organized as follows:

$$EE \text{ (Daily Energy Expenditure for each activity)} = (N * D * MET \text{ value}) / 365,$$

where

N is the number of times the agent engages in each sport during the whole year;

D is the duration or hours the respondent usually spends on each activity;

MET value is the corresponding expenditure multiple that is represented in the appendix (table 1).

In addition, because the survey designs the questionnaire as the “the number of times take this exercise in the past three month”, we need to multiply 4 when we calculate the EE. Moreover, the questionnaire divided “time spent on exercise” into four categories: 1 to 15 minutes, 16 to 30 minutes, 31 to 60 minutes and more than 1 hour. For those multiple categorical data, average values of exercise duration are used during the calculation. The CCHS refers to the method of National Population Health Survey (NPHS): 13 minutes or .2167 hour, 23 minutes or .3833 hour, 45 minutes or .75 hour, 60 minutes or 1 hour are allocated into each category of “time spent on

exercise” respectively, for the convenience of calculation. There are also several confounding categories, like “not applicable (do not participate in this exercise)”, “do not know”, “refusal” and “not stated”. We could have deleted the data of those responders who do not answer the required questions (“do not know”, “refusal” and “not stated”), while kept the “not applicable” category data. The survey includes 26 kinds of exercise. It is very likely for agents to miss one or several questions. If we delete all of this responder’s related information, we may lose a lot of significant information. Only by the tool of STATA cannot strip out some information individually from a respondent's overall information. We follow the recommendation of the CCHS and assign a value zero to those vague answers. This means that CCHS treats those who didn’t offer clear answers as people who didn’t participate in this exercise. We think that this approach is reasonable to some extent, since their refusal to answer could indicate that they might not familiar with several specific exercises.

3.2 Gini Coefficient and Lorenz Curve

Initially, in order to study the national income distribution situation across the population, the American statistician Lorenz proposed a famous Lorenz Curve. Gradually, it is more and more widely used in the analysis of inequality distribution. In the field of health economics, Wagstaff et al. (1991) mention that Le Grand (1985) is the first to adapt the true Lorenz Curve in health inequality. In our paper, we extend the use of Lorenz Curve to the area of physical activities in order to study the inequalities distribution.

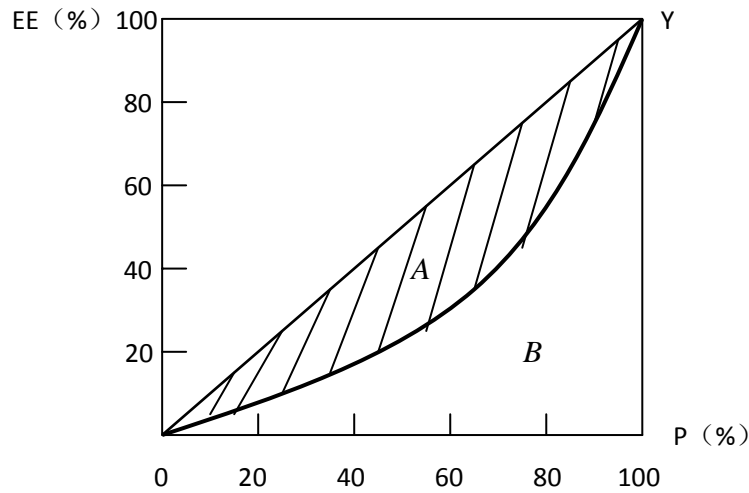


Figure 1: Energy Expenditure Lorenz Curve

This graph plots the energy expenditure Lorenz curve. The abscissa represents the cumulative population ranked by average daily energy expenditure. From left to right, people are arranged in ascending order according to the daily energy consumption. The ordinate demonstrates the cumulative energy cost. The Lorenz curve is the line in the middle of the graph, which indicates the relationship between the cumulative proportion of population (from the leftmost person who takes the least exercise to the rightmost one who spends most daily energy) against the cumulative proportion of energy expenditure. It could be a straight line or a curve. If all the points fall on the diagonal line, it coincides with the equality line or egalitarian line (Zhang and Wang, 2004) and the whole population takes the equal exercise intensity. When the Lorenz curve is concave, each corresponding points lie below the diagonal line. Specifically, we can analyze two points (20,10) and (70,40) as two examples. (20,10) states that 20 percent of the population with the least exercise corresponds to 10 percent of exercise energy consumption. Similarly, (70,40) indicates that 40 percent of the cumulative energy expenditure is shared by 70 percent of people. If there is no inequality distribution, the variables at both horizontal and vertical coordinates should

increase at the same growth rate. Therefore, it is the unequal distribution of exercise expenditure that twists the Lorenz curve.

The study concerning the degree of curvature of the Lorenz curve is meaningful, and reflective of the degree of inequality. The farther the curve is away from the diagonal, the greater the curved arc is, and correspondingly more serious the inequality is. The shaded area between the diagonal and the Lorenz curve in the figure represents the inequality area (area A). Obviously, larger area means larger inequality. Gini coefficient is related to this area A .

$$\text{Gini Coefficient} = \frac{A}{A+B} = 2A$$

Where, area B is the area below the Lorenz curve.

Because $A+B$ is equal to $\frac{1}{2}$, Gini coefficient can also be expressed as $2A$. Here, A and B are both percentage values. When area A is equal to zero, Gini coefficient is also equivalent to zero. The population shows mean absolute deviation. The relevant Lorenz curve coincides with the diagonal line and becomes the line of perfect equality. When area B equals zero, the value of Gini coefficient is one. The Lorenz curve coincides with the horizontal and vertical coordinates, showing the shape of a right angle. It reflects the perfect inequality. Those are two extreme cases. Normally, the value of the Gini coefficient is between zero and one. A larger value is combined with larger inequality, vice versa.

3.3 Concentration Index and Concentration Curve

Although the Gini coefficient and the Lorenz curve have been widely used

because of their simple operations, they still have many defects. Both of them cannot respond to changes in energy expenditures of different classes. For example, when people in class I (with least amount of exercise) start taking more physical activities and people in class IV (with largest amount of exercise) reduce the exercise time, the inequality will become smaller. However, Lorenz curve cannot identify such kind of change. As far as the purpose of our work is concerned, the most serious problem is that the Gini coefficient and the Lorenz curve only state one single aspect of the problem (inequality in physical activity), and they cannot help to explain the socioeconomic situation. The most important question is that whether the physical activities have the distinct distribution between the rich and the poor, i.e. how the inequality changes among the different socioeconomic status.

Wagstaff et al. (1991) compare several methods to measure health inequality. They state that the concentration index (*CI*) and the slope of inequality are the most accurate ways to analyze the inequality in health. The latter one has the similar feature as the former. Wagstaff et al. (1991) gave *CI* a high evaluation: *CI* could demonstrate how the target variable distributes among the various socioeconomic status populations, taking the whole population into account. In addition, *CI* is also sensitive to the change in different socioeconomic status. Based on the Yitzhaki's (1983) extension of Gini coefficient, Wagstaff (2002) conducts a series of application and extension for *CI*.

In our paper, we also choose the *CI* as our main analysis tool. Because our purpose is to study socioeconomic inequalities in physical activities participation, we

have to pick up the reasonable variables. In the first part of section three, we have described the average daily energy expenditure, which can be a good measurement of physical activities from frequency, duration and intensity aspects. In terms of the socioeconomic status, there are several relevant variables, such as household income, education level and occupation. We can easily understand that socioeconomic status is a complicated concept, which is affected by income, family background, credit, position, education etc. Here, we select household income as an alternative to measure socio-position. As Zhang and Wang (2004) mentioned in their work, income always shows the positive relationship with occupation, position, education and so on. Other indexes, like education and occupation, tend to be stable among the adults and would mask the socioeconomic variation in exercise. Income index is more sensitive to socioeconomic status and can be a good alternative of socioeconomic position.

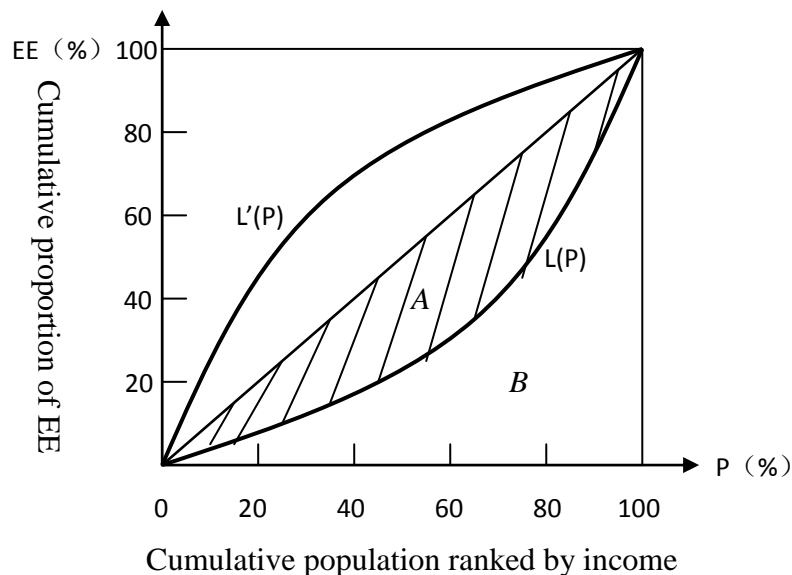


Figure 2: Energy Expenditure Concentration Curve

Figure 2 depicts the energy expenditure concentration curve. Here, we use the

artificial example for the convenience of explanation. The horizontal axis (P) represents the cumulative proportion of population arrangement ranked by income. People are arranged in an ascending order, according to the level of income (starting from the poorest one at the far left to the richest person at the far right). The vertical axis (EE) indicates cumulative proportion of daily energy expenditure. $L(P)$ and $L'(P)$ in the graph represent the concentration curve (*CI* curve). The amount of energy expenditure (physical activity) is added according to the corresponding information of the social class. Unlike the Lorenz curve, the *CI* curve could be above or below the diagonal line, i.e. $L'(P)$ and $L(P)$ respectively. We can discuss, for example, the point (40, 20) on the $L(P)$ firstly. This point means that the 40 percent of the population with the least income only takes 20 percent of the physical activities. The rich participate in more exercise than the poor. The physical activity favors the rich. According to Zhang and Wang's (2004) description, this kind of line can be called a "regressive curve". Then, with regards to the above line $L'(P)$, we pick up (40, 60) as an example. The 40 percent of the people who stay at the lowest social class actively take 60 percent of the physical sports. The rich seems to be lazier. Physical activities are pro-poor. The line above can be called "progressive curve". Compared to the work of Zhang and Wang (2004), we have the different position of those two lines. Their study is related to the obesity and socioeconomic inequality. "Obesity" is a bad variable compared to "daily energy expenditure". Thus, in their work, the line below the diagonal is pro-poor (progressive curve), while the line above the diagonal is pro-rich (regressive curve). In principle, the two applications are the same. *CI* is the

twice area as the shaded area A . As mentioned in the above, area A is the area between the diagonal and the CI curve. Or, we can also say CI is the area of one minus twice area B . Area B is embraced by the CI curve and two coordinates. The formula is as follows:

$$CI = 1 - 2 \int_0^1 L(P) dp$$

The value of the CI is between -1 and 1. In our case, for the “regressive curve” which is pro-rich, CI is positive. When the CI is larger, the distance between the CI curve and the diagonal line will become larger. Correspondingly, the socioeconomic inequality is more serious. When CI shows the negative value, the result is pro-poor (progressive curve).

Here we also introduce the extended CI and its transformation. At the end, we can state that this concentration index also gives the poor people more weights. Therefore, it is more sensitive to socioeconomic inequality. Kakwani, Wagstaff and van Doorsaer (1997) proposed another way to express the CI . The formula is presented below (we explain this formula under our physical activity case):

$$CI = \frac{2}{n\mu} \sum_{i=1}^n y_i R_i - 1$$

Where, n is the number of all the responders; μ is the mean value of the daily energy expenditure; y_i is the relevant energy indicator of the person i ; R_i is the rank of individual’s position in socioeconomic status (income position). After several changes for this, we can get a new formula for the CI :

$$\begin{aligned}
CI &= \frac{2}{n\mu} \sum_{i=1}^n y_i R_i - 1 \\
&= \frac{2}{n\mu} \sum_{i=1}^n y_i R_i + [1 - 2(\sum_i \frac{y_i}{n\mu})] \\
&= 1 - \frac{2}{n\mu} \sum_{i=1}^n y_i (1 - R_i)
\end{aligned}$$

From this equation, we can see that $\frac{y_i}{n\mu}$ is the individual's share of physical activity participant. When the person i is at the left most poorest position, R_i is close to 0. Then a value 2 will be assigned to person i as a reasonable weight while calculating CI . If the agent i stands at the right most richest level, the related R_i value will be close to one. He almost gets zero of weight.

3.4 Achievement Index

Before we introduce the achievement index, we provide the formula in the first place and then make further discussion. Achievement index is denoted by AI :

$$AI = \mu[1 - CI]$$

From the formula above, we can clearly see that the achievement index includes analysis of both the socioeconomic inequality, as represented by the CI and mean value of energy expenditure (μ). This is a more powerful method to measure the inequality in physical activity of participants. Let us discuss this in detail. In our case, the CI is positive when the exercise is pro-rich. Meanwhile, mean value μ is always larger than zero. Therefore, a smaller AI may tell a more serious situation and we prefer a larger achievement index. Suppose that for one group of population, they have larger CI , then the AI will get smaller, which may state that there exists

the socioeconomic inequality of physical activity among them. At the same time, if they still have very low mean value of exercise indicator, the whole exercise situation for this group will become more serious and the exercise condition for the poor will be much worse. However, with the same high CI , if the group obtains a larger value of mean (μ), then AI will become relatively larger. This means that although they have inequality to some extent, a relatively high mean value also lifts the situation of the poor, whereby at least it is not too bad from the aspect of overall exercise amount.

Thus, in our paper, because of its comprehensiveness, after we study the concentration index, we still employ the achievement index to assist the analysis. In addition, we can also analyze the different effects on the poor and the rich with the various weights given to them. We can make some modification to the formula above:

$$\begin{aligned}
AI &= \mu[1 - CI] \\
&= \mu - \mu CI \\
&= \mu - \mu \left[1 - \frac{2}{n\mu} \sum_{i=1}^n y_i(1 - R_i) \right] \\
&= \mu - \mu + \frac{2}{n} \sum_{i=1}^n y_i(1 - R_i) \\
&= \frac{2}{n} \sum_{i=1}^n y_i(1 - R_i)
\end{aligned}$$

Obviously, the achievement index can be seen as a sum of weighted exercise situation for each individual. The poorer person earns a larger weight, while the rich one obtains smaller weight.

3.5 Decomposition of Concentration Index

Relying solely on general concentration index does not reflect the deeper

problem. When we have already answered the overall inequality through the concentration index, we still want to know which part of exercise contributes most to the overall inequality and how different subgroups contribute to overall *CI*. Thus, we could carry on the study into more disaggregated stage. In our work, physical activities consist of 26 different representative exercises. When we calculate the average daily energy expenditure, we firstly compute each energy cost value of the corresponding exercise and then we add each item together. Then in the first place, we can decompose concentration index by various physical activities. Here, we use the decomposition method proposed by Clarke et al. (2003). The basic formulas are as follows:

$$E_i = \sum_{k=1}^K \phi_{ki}$$

There are K kinds of different activities (in our survey, K is equal to 26). Lowercase k represents every exercise. ϕ_{ki} is daily energy expenditure of sport k for person i . E_i adds all of the items together and represents daily exercise energy consumption for individual i .

$$\mu_E = \frac{1}{n} \sum_{i=1}^n E_i$$

Here, μ_E calculates the average value of daily activity energy for the entire population.

$$\mu_{\phi_k} = \frac{1}{n} \sum_{i=1}^n \phi_{ki}$$

For each activity k , μ_{ϕ_k} stands for mean value of daily activity k energy consumption of all people.

$$CI = \sum_{k=1}^K \frac{\mu_{\phi k}}{\mu_E} CI_k$$

The overall concentration index is denoted by CI . CI_k stands for each corresponding concentration index of various activities. CI is a weighted sum of CI_k . The relevant weight is $\frac{\mu_{\phi k}}{\mu_E}$. On one hand, μ_E is the mean value of the entire exercise components. On the other hand, it is also a sum of all average exercise energy denoted by $\mu_{\phi k}$. Therefore, $\frac{\mu_{\phi k}}{\mu_E}$ reflects the specific energy cost proportion of activity k in all K . At the same time, we can also compute the contribution of exercise k to overall CI , as well as the relative contribution of overall CI .

$$\text{Contribution of } k = \frac{\mu_{\phi k}}{\mu_E} CI_k$$

$$\text{Relative contribution of } k = \frac{\mu_{\phi k} CI_k}{CI} * 100\%$$

From those several indexes above, we can clearly explain the contribution effect of each activity to the overall inequality. It offers us an evident comparison among each other. Some institutions could also focus on several certain exercises with large value so as to put forward some effective ways to relieve the inequality.

Besides the socioeconomic status, there are also other related demographic factors which affect the inequality in participant's physical activity, such as gender, age, ethnicity and immigration year. We cannot directly include any such kind of factor into CI , because it includes all of the demographic factors together into the socioeconomic status. However, we can stratify the population according to a specific characteristic and then analyze the CI among the corresponding subgroups. The

general CI also includes the understanding of inequality “between groups” and “within group”. This decomposition method is presented by Clarke et al. (2003).

$$CI = CI^W + CI^B + CI^T,$$

where $CI^W = \sum_{j=1}^J \theta_j \frac{\mu_j}{\mu} CI_j$. Meanwhile, CI represents the overall concentration index.

CI is the sum of CI within subgroup (CI^W), CI between subgroup (CI^B) and transvariation term (CI^T). When we conduct the decomposition work by population, we start from the relative simple calculation of the overall CI . Then we work out each CI_j for corresponding subgroup. The index J represents the total number of the target population groups. CI^W is defined above. μ_j is mean value of energy expenditure for relevant population. μ stands for the average for the whole responders. θ_j is the related population share for each subgroup. We can take “gender” as an example. The entire population can be divided into two groups, i.e. male group and female group. In this case, J is equal to 2; θ_1 is the male population share and θ_2 is the female population share. When we calculate CI between groups, we set the related original average subgroup value to each person within this group. Then we calculate the overall CI based on two groups of mean values, which is equal to CI^B . Subsequently, we also get the transvariation term by subtracting CI^B and CI^W from CI . If CI^B is larger than CI^W , the main inequality is reflected between the subgroups. In this regards, as for our example, a larger CI^B states that the inequality between the female and the male is the more important. If CI^W has a larger value, the socioeconomic exercise inequality within groups has a greater influence on the inequality. The last term (CI^T) is the transvariation term or a residual, which is

explained by two things: one is from the overlapping part between the densities of EE of two groups; the other one is from the difference in the ranks of an individual within his group and in the overall population.

3.6 Decomposition of Achievement Index

From the introduction above, we can also think about taking the identical decomposition for achievement index, due to their similar features. The achievement index (AI) can be obtained by $AI = \mu(1 - CI)$. In our paper, we disaggregate it from two dimensions: by categories of physical activities; and by population subgroups. As for the former case, we calculate each achievement index on the basis of the mean value of each exercise and each concentration index of every activity, i.e. $AI_k = \mu_k(1 - CI_k)$. Then through the composite analysis, we can compare and discuss the socioeconomic inequality of each activity.

With respect to the second case, we follow the method that Makdissi et al. (2013) mentioned in their work concerning health inequality. Firstly, we stratify the whole population by various characteristics, such as gender, age, ethnicity, and immigration year. The energy expenditure achievement can be composed in the following way:

$$AI = \sum_{j=1}^J \theta_j AI_j + AI^T$$

Here, J is the number of the population subgroups (for gender groups, $J = 2$). θ_j is the corresponding population share; AI_j is the relevant achievement index of each group. AI^T is the transvariation term. According to what Makdissi et al. (2013) explain, the residual term AI^T comes from the distinction between different ranks of

individual within specific subgroups and in the entire population.

4. Empirical Analysis

In this part, we will describe our empirical results of physical activities inequality across the various socioeconomic statuses in Canada. Our main target is to analyze the concentration index and the achievement index.

4.1 Data and Variables

We use the data collected by the Canadian Community Health Survey (CCHS), 2012. The CCHS is a cross-sectional survey that provides the comprehensive information related to health condition (various health indicators and regular living habits), public health service status, and related personal details. The survey contains the relevant data from those who are 12 or more than 12 years old in ten provinces and three territories. The survey excludes the people who live on reserves and other Aboriginal settlements, Canadian full-time soldiers, institutionalized population and people who live in the Quebec health regions of R égion du Nunavik and R égion des Terres-Cries-de-la-Baie-James. The survey's coverage is very large, accounting for more than 97% of the target population. We extract several useful statistics from this large data set, such as "total household income from all sources", "sex", "age", "immigrant", "length in Canada since immigration", and some related data of physical activities concerning both frequency and duration. We have taken into account statistical weights in all estimations.

As for income variable, we follow the advice of many related study on health

inequality and choose the household income as our target because of its stable value within a year. The data includes five categories: “no income or less than \$20,000”; “\$20,000 to \$39,999”; “\$40,000 to \$59,999”; “\$60,000 to \$79,999” and “\$80,000 or more” (refer to table 2 “Household Income from All Sources” in the appendix). Around 33.3 percent of the people have \$80,000 or more revenue every year. When we work out for *CI* and *AI*, we should pay attention to the responders who didn’t answer the required answers, such as the one who chooses the category of “do not know”, “refusal” and “not stated”. Thus, before we start to calculate the values, we drop such responders’ information according to income. Fortunately, there are only 60 people out of 61707 people in total who choose “not stated” and we still keep most of the observations.

With regard to the age group, we follow the classification from the CDC (Centers for Disease Control and Prevention). We divided the responders into three categories: 12 to 17 years of age (children), 18 to 64 years old (adult) and 65 years of age or older (older adult). As for the length in Canada after immigration, we choose 10 years as a threshold to divide the population into two groups: one with less than 10 years, and the other one with 10 years or more.

4.2 Physical Activity Index

The objective of our work is to study the socioeconomic inequality in physical activity. We cannot simply select a variable from the ready-made data, like frequency (how many times you did this exercise in the last three months) or duration (how long

do you spend on this activity). It is necessary to include each MET into the study. Thus, we introduce daily energy expenditure to describe the individual exercise situation better. Firstly, we calculate everyone's daily energy cost for each specific exercise. In that way, we obtain everyone's 26 statistics for corresponding 26 kinds of sports. 26 is the number of the physical activities in the survey, including swimming, bicycling, hockey, dance etc. Then, from the perspective of each person, we add those data together so as to get the individual's total daily energy consumption (TDEE) on exercise.

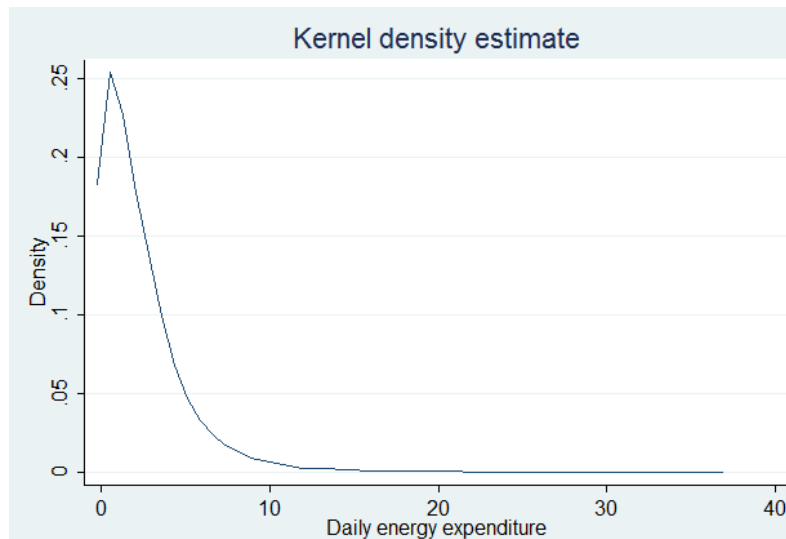


Figure 3: Density Graph of Daily Energy Expenditure

Table 1III represents the estimated sample mean of daily energy expenditure of 2.3540. This value at the nearly leftmost part of the distribution of overall daily energy expenditure. According to Figure 3, the density curve indicates a very broad distribution, with a long tail in its right part. Most of the Canadians stay at the inactive or moderate position, while there is still a relatively small percentage (28.81%) of people who are into active exercise. Table 3 shows different exercise intensity of

Canadians and presents the physical activity index.

We also calculate the Gini coefficient of energy expenditure (0.5216) and construct the corresponding Lorenz curve (Figure 4). Like what we discuss above, the Gini value larger than 0.5 indicates a large gap among the activities distribution. In addition to this, we still need to study the relationship between socioeconomic dimension and exercise inequality. Meanwhile, it is vital to include other related factors, which affect the exercise amount into our analysis.

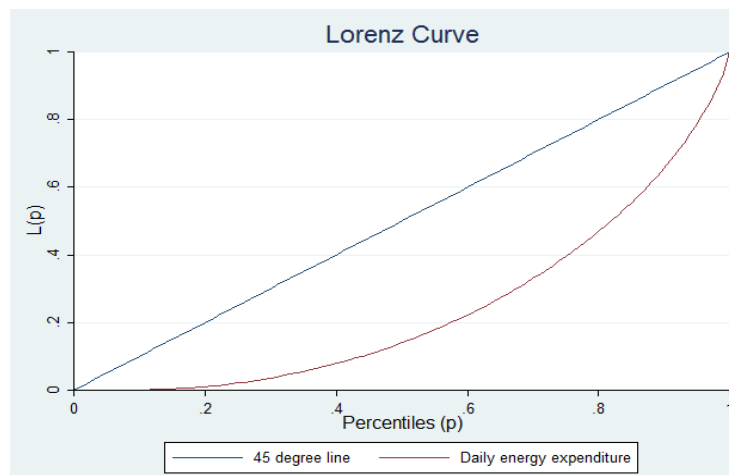


Figure 4: Lorenz Curve of Daily Energy Expenditure

4.3 Decomposition by Categories of Physical Activities

In the following part, we will discuss the core socioeconomic inequality with the assistance of concentration index and achievement index. As what we introduce in the text above, the concentration index is obtained by twice the area between the diagonal and concentration curve, which focuses on the measurement of the overall inequality in physical activities by income.

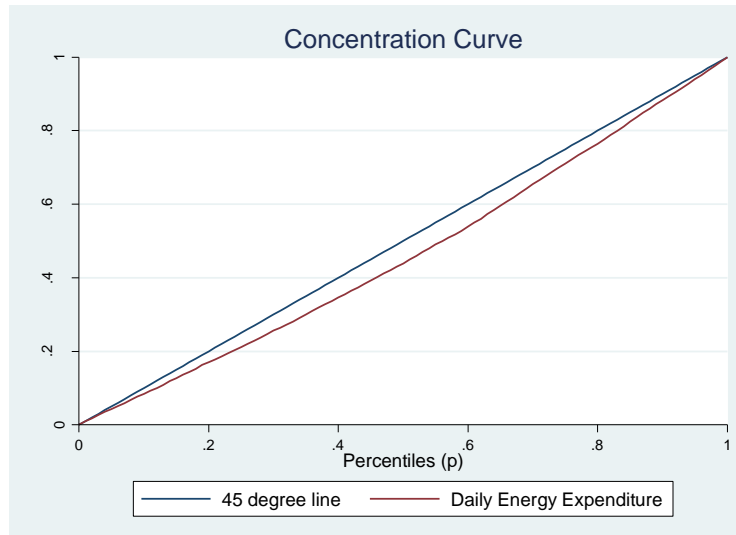


Figure 5: Concentration Curve of Overall Activities

Figure 4 presents the pure inequality of physical activities in the population. Figure 5 introduces the individual's household income and demonstrates the socioeconomic inequality in exercise. Comparing Figure 4 and Figure 5, we can easily find that the CI (0.0663) and the concentration curve (Figure 5) display less inequality than the Gini (0.5216) and the Lorenz curve (Figure 4). The distance between the concentration curve and the diagonal is relatively smaller than that for Lorenz curve. This is a good situation, and we infer that there should be a large overlapping part in energy expenditure density between the poor group and the rich group. We combine the 3 poorest groups together (people with income equal to or lower than \$59999), and also combine 2 richest groups together (people with income equal to or more than \$60000). Their density distributions are shown in Figure 6.

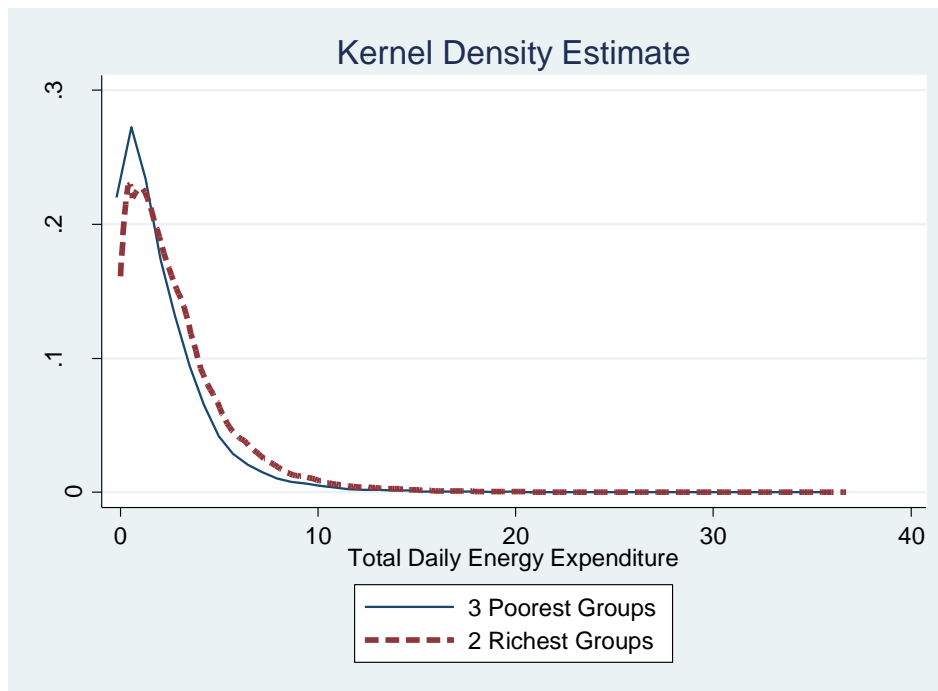


Figure 6: Multiple Densities Graph of the Rich and Poor

On Figure 6, the thin solid line stands for the distribution of the 3 poorest groups, and the thick dash line is the density distribution from the 2 richest groups. The latter rich curve stays a bit right the poor curve, which comes from the socioeconomic inequality of physical activities among overall population. Meanwhile, we also find that those two densities overlap a lot at the energy expenditure between 0 and 10. Therefore, we can conclude that, like the poor group, a large part of rich people stay at relative inactive or moderate exercise situations. The physical activities inequality among the various socioeconomic statuses exists but it is not so obvious.

Table 4 presents the concentration index, contribution of each activities and relative contribution, respectively. *CI* is decomposed by various categories of physical activities. From this table, we can clearly see the inequality influence of each exercise. Firstly, let us pay attention to the *CI* separately. Most of the statistics are positive. Only three of them are negative. The smaller the estimate is, the better is the

exercise distribution. When we add income as one significant dimension, we get the overall *CI* of 0.0663. Compared to the Gini result, we obtained in the text above, the activity inequality seems getting better. However, there still exists some physical activity inequality among Canadians. The three negative values comes from “walk for exercise (-0.0271)”, “bowling (-0.0438)” and “transportation-walking (-0.0348)”. Those three values alone may not explain the real story of the exercise situation. After we add the corresponding weights as in the formula of relative contribution of k , we get the relative contribution. “Walk for exercise (-11.73%)” becomes more outstanding, while “bowling (-0.22%)” and “transportation-walking (-1.71%)” decrease their influence. Now we can speculate that those people who are at the poor position prefer to choose walk as their main exercise. In addition, still from the concentration index, ice hockey (0.3324), golfing (0.2475), ice skating (0.2311) and tennis (0.2161) account for the top four contributors with positive value. Meanwhile, the activities such as downhill skiing, baseball, weight-training, volleyball and soccer also present relatively large numbers. A larger number for the *CI* means more inequality in those activities by income status, which reflects the degree of inequality in exercise distribution related to affluence. Those sports with high *CI*s always need expensive sports equipment, specialized exercise places, and sometimes professional counseling coach. Like ice hockey, and ice skating, except for outdoor place in winter, those sports need to be in the indoor skating rink. Golf’s requirement on sites and equipment are obviously higher. Tennis is also in the same case. In addition, we also find that for those team sports, they tend to have higher values of the *CI*. We could

explain that team sports games often require a professional coach and specialized sports venues. Meanwhile, team sports, which are in the form of clubs, may require a membership fee. For the people with average or low income, those kinds of activities are not the favorite choices. Low-income people tend to choose some simple and easy operation sports, such as bicycles, social dance, fishing and the like. As for home exercise and gardening or yard work, although they have the positive values, they do not show a large socioeconomic inequality.

As what we discussed before, the achievement index also takes the mean value into consideration. When we move to the study of the corresponding achievement index, we can see more clearly about the real inequality situation for exercise in Canada. Achievement index is obtained by $AI = \mu(1 - CI)$. By definition, this indicator is always positive. We can measure the inequality through the comparison of the index size. Because the energy expenditure of exercise is something beneficial to our health, we always prefer a large mean value μ . Intuitively, the larger achievement index is, the better is the exercise distribution by socioeconomic status. This logical thinking is right. When the CI has a positive value (regressive), the larger CI which means the larger inequality will cut down the AI value. A larger μ indicates that the average participation in physical exercise is high. If it is combined with a large CI , the AI value will be dragged down. Conversely, if we have a small CI , it increases the value of AI , compared to μ . Therefore, the achievement index is good indicator and we prefer a larger value of AI . Table 6 displays the relevant CI , mean value and achievement index of each activity. From

the mean values solely, “walk for exercise” account for the largest part. This mean value (0.6750) is larger than twice the second one “gardening or yard work” (0.2904). We can say that most of Canadians choose walk as their daily exercise, because from the overall mean value, Canadian’s energy expenditure on walking is high enough. Thus, the corresponding *AI* value is also very large. Several other large mean values are “gardening or yard work”, “bicycling”, “home exercise”, and “jogging or running”. Canadians spend relatively more energy on those activities, with more frequency and/or larger duration.

When we discuss the *CI* separately, we find that ice hockey, ice skating, golfing and tennis could be the chief culprits for the entire inequality. At the same time, we carry on the comprehensive analysis including the mean value approach. From table 6, the average daily energy expenditures of those activities are also very small. Especially for tennis sport, it has a pretty big *CI* but a very small mean value. This kind of combination makes the achievement index even smaller. As a result, the influence of those kinds of sports on socioeconomic inequality is magnified. Now we conduct a deeper analysis for the achievement index. Table 6 represents the achievement indices decomposed by various activities. Figure 7 depicts the corresponding achievement indices for different excise. From Table 6 and Figure 7, we can state that the best five performing physical activities are “walk for exercise (0.6933)”, “gardening or yard work (0.2796)”, “jogging or running (0.2214)”, “home exercise (0.2077)”, “bicycling (0.1126)”. All of them have several common features: they do not require the special venues; they can easily accessible; There is no need for

expensive equipment; all the exercise expenses are close to zero. Although several of these physical activities have positive (regressive) *CI* values, their relatively large average daily energy expenditure shows that those activities are relatively equal across various socioeconomic status.

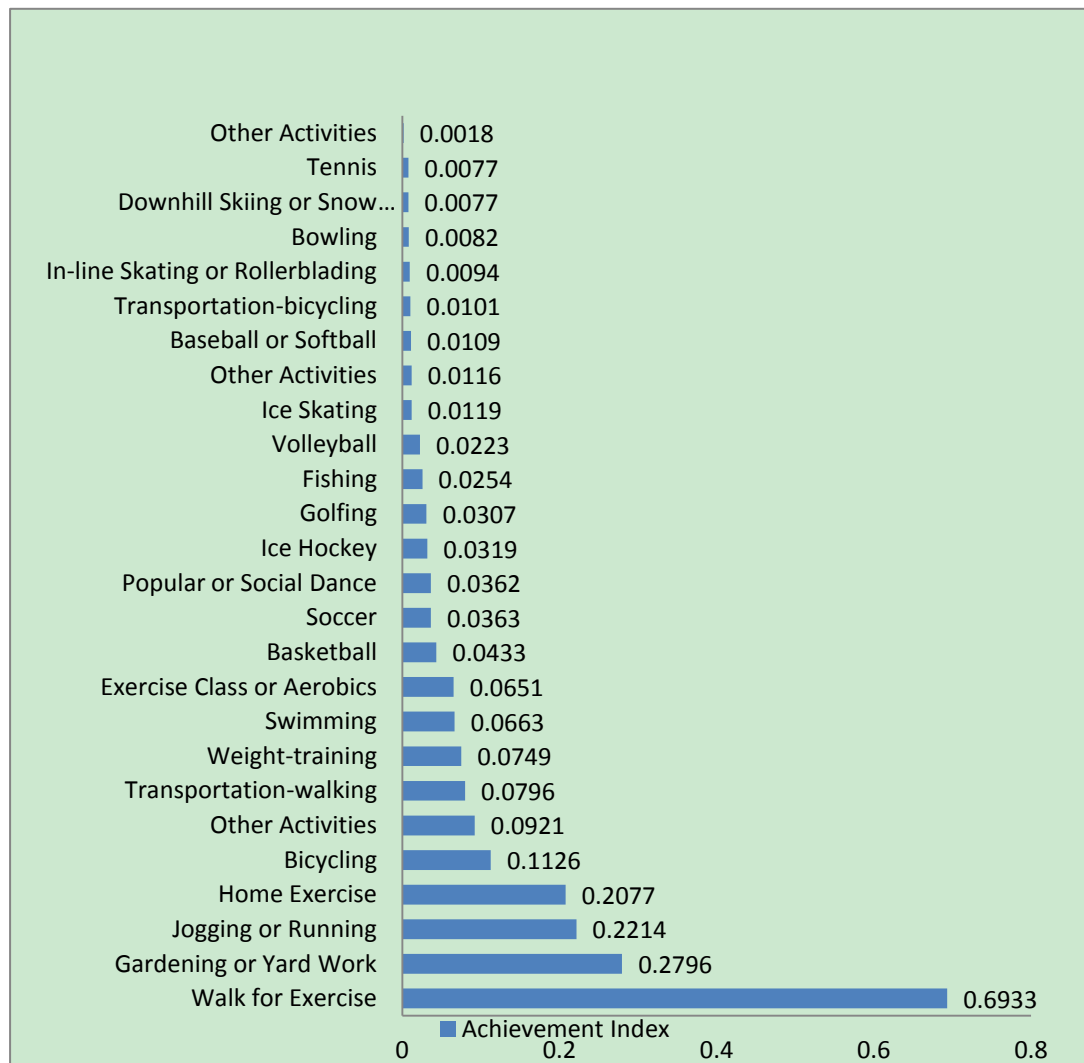


Figure 7: Achievement Index Decomposed by Activities

As for several data with bad performance, like “tennis”, “bowling”, “downhill skiing or snowboarding” and “in-line skating or rollerblading”, they have very small values, which are all less than 0.01. Bowling value is very interesting here. It has a negative value of the *CI*, while it also has very small values of both mean and the

achievement index. The relevant MET for bowling is 2. We can explain that people with low income could prefer bowling. However, because its low energy consumption, this might not be a very popular exercise among the entire population. Tennis also has the same problem. The *CI* for tennis is high, while its mean value is much smaller. Tennis sport could be much more popular among the rich people because of the special requirement for places and coach, as well as expensive rackets. Another interesting activity is jogging or running. Although it also has a high value of *CI* (0.1780), with a high mean value, it obtains a satisfactory achievement index. Canadians with larger income may take more jogging or running compared to the poor. However, due to its easy access, Canadians with lower income also participate in these two exercises. Meanwhile, with a high value of MET (9.5), jogging or running could be an effective way for Canadians to prevent obesity.

4.4 Decomposition by Population Groups

Besides the income position, there are other factors that could affect the physical activity participation, such as gender, age, education and so on. However, we cannot technically include those factors in our measurement because of the limitation to adopt two-dimension *CI*. We can still involve the analysis of those demographic characteristics. The best way is to stratify the entire population according to a target characteristic and then carry on the research on the socioeconomic inequality in the specific groups. In our paper, we select 4 related factors: gender (male and female), immigrant (yes or no), length of years after immigration (less than 10 or more) and age

(12-17, 18-64 ,64+).

As for the concentration index, $CI = CI^W + CI^B + CI^T$. In order to calculate CI^W , we have to work out each CI for each subgroup. Then we also need to assign the corresponding subgroup population share and mean value ratio to each CI . CI^W can state the income related effect on inequality. In terms of CI^B , we set the value for each person to the average of the subgroup and then calculate the overall CI with those stepped data.

Table 5 lists a series of tables concerning the decomposition of concentration index by four characteristics. Let us analyze them one by one. Firstly, we decompose CI by gender into two groups: male group and female group. As for each separate CI , $CI_{male}(0.0722)$ is larger than $CI_{female}(0.0675)$. Both of them have socioeconomic inequality within group to some extent. The difference between two seems to be not very significant. This intuition could be correct. CI^W (0.0698) for them accounts for 95.74% of the total CI , compared to the smaller value of $CI^B(0.0049)$ which only explains 6.76% of the total. Therefore, we can conclude that inequality between men and women is not so obvious. Most of the energy expenditure inequality is reflected on the socioeconomic position within group. In the case of immigration status, CI^B (0.0019) presents a much smaller value, which means that there is almost no difference among various immigration statuses. This result is somewhat consistent with Zhang and Wang (2004). They state that minority groups do not have a higher SES inequality in obesity. Since people have their own stable diet and exercise habits, they will still keep those living styles after their immigration. A large CI^W (0.0712)

value accounts for 102.74%, leaving a negative value for transvariation term. Here, income inequality in physical activities within group again shows its larger effect. We also strip the immigrants out and consider the length of years after immigration. $CI_{\text{less than 10}}$ is negative while $CI_{\text{more than 10}}$ is positive. During the first several years after immigration, the socioeconomic inequality in exercise does not show up among the new immigrants. They are all at the stage of adaptation from various aspects. Later, when they stay in Canada long enough, they gradually integrate into the local community. Their physical exercise situation is getting closer to native Canadians. $CI^B(-0.0060)$ with -19.3% can explain the relative large inequality difference between less than 10 years and more than 10 years' immigration. There is another interesting point that the transvariation term (-29.41%) here are very high. We draw a kernel estimation of the densities of energy expenditures of people with less than 10 years and with more than 10 years' immigration on the same graph. From the Figure 8, we can infer that this high CI^T comes from the large overlap of the densities of the two groups. Immigrants with less and more than 10 years' immigration may keep relative same exercise situation.

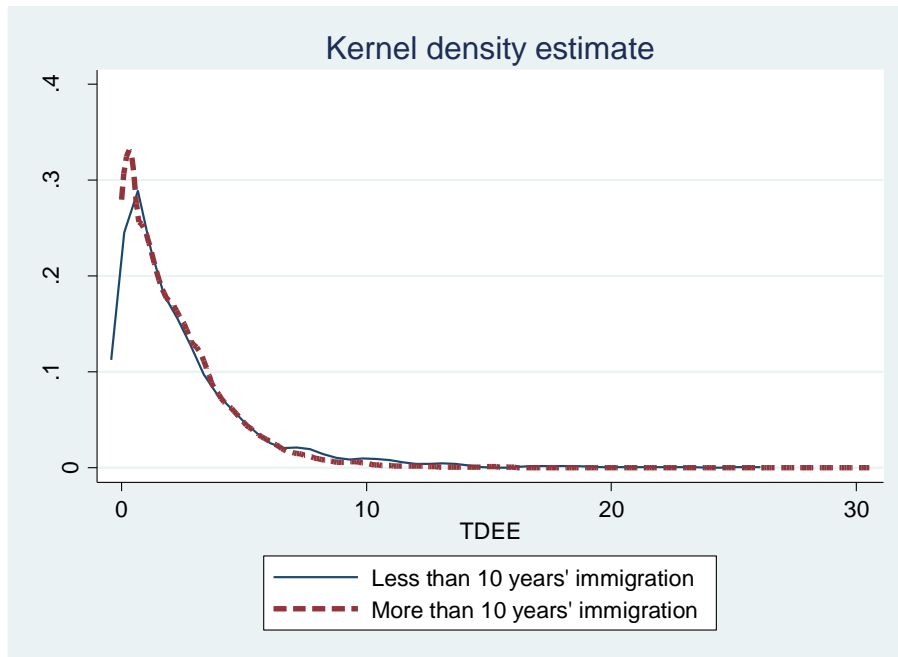


Figure 8: Multiple Densities Graph of the Two Immigration Groups

We also decompose CI into three groups by age: children (between 12 to 17 years old), adult (between 18 to 64 years old) and older adult (65 years old or older). Firstly, we discuss the CI for each subgroup separately. All of them have positive values. The older adult group gets the highest CI (0.0856) value. Adult group (0.0545) is in the middle and child group (0.0215) has the smallest CI . Socioeconomic inequality in sports participant is more obvious among the older adults. From those three values, it seems that inequality has the different distribution among the three groups, because of the relative large difference among three CI s. Just as expected, CI^B presents a large value (0.0215) and accounts for 29.49% of the total inequality. It is simple to understand. Since those three age groups have distinct income structure and consumption habit, as well as different exercise types and intensity, their income inequality in activities among those three is obviously different. CI^W still keep a higher value (0.0553), which takes up 75.86% of contribution.

Although it is still large, it is reduced a lot compared the other CI^W we discussed before.

Then we focus on the achievement index. Table 7 presents 4 groups of decomposition results. For the gender group, both of the male group and female group present almost the identical data. Men do a little better than women, because of higher mean value than women. We could infer that the activities men take part in have more energy expenditure intensity than women. This is also reasonable. As for the immigration status, there is less difference between yes-group (2.1177) and no-group (2.2080). This result is also consistent with the result from the concentration index. An ethnic background may not have a large effect on the activity inequality. Another interesting point is related to the length in Canada since immigration. Compared to those with more than 10 years' immigration, people who immigrate into Canada less than 10 years have a larger value of achievement index due to their negative CI and a larger mean value. This indicates that, as they stay in Canada longer and longer, the average amount of exercise for those immigrants is falling down. Meanwhile, the physical activity inequality by socioeconomic status gradually widens. With regard to three categories of the age group, there are relative large differences among those three. Children at the age between 12 to 17 years old have a largest achievement value. If we look at their CI and mean value, they also have the smallest CI and the largest mean value. The result suggests that children have the smallest exercise inequality and the maximum exercise intensity. Conversely, for the elder, they cut down their exercise and also, they begin to show a larger activity inequality across

various income positions.

5. Conclusion

The study of socioeconomic exercise inequality is significant in health economics. We use the Gini coefficient to test the degree of inequality in exercise independently. Activities have a large distribution with a relatively large inequality. Most of Canadians stay at the lower position due to the inactive and moderate exercise intensity. Then we also need to employ the income factor. We select concentration index (*CI*) and achievement index (*AI*) as our main tools to measure the physical activity inequality across various socioeconomic status. Decomposition work makes our analysis into a disaggregated level. Based on the method offered by Clarke and Makdissi, we decompose the *CI* and *AI* by different physical activities. The results indicate that the poor tend to choose a single exercise with public free exercise place, like walking, bicycling, gardening and home exercise. Especially, the *AI* value of walking for exercise is very high. However, tennis, ice hockey, and golfing have smaller *AI* value. We can infer that those Canadians' with high income prefer the club activities and team sports. On the whole, the poor take less physical activities and less sports types than the rich, because of their less stronger awareness of exercise than the rich, to some extent. When we decompose *CI* and *AI* by demographic factors, we find that the socioeconomic health inequalities within group are the main contributors. All of the indices between groups are not very high. Among them, the length of time since immigration and the age present the relatively large

values, but still less than the indices within groups. That indicates that the exercise inequality between immigration years less than 10 and more than 10 is more obvious. Meanwhile, the socioeconomic inequalities in physical activity between age groups are also more outstanding.

6. Appendix

Table 11: MET Table for Physical Activities

Average Daily Energy Expenditure	MET
Walk for Exercise	3
Gardening or Yard Work	3
Swimming	3
Bicycling	4
Popular or Social Dance	3
Home Exercise	3
Ice Hockey	6
Ice Skating	4
In-line Skating or Rollerblading	5
Jogging or Running	9.5
Golfing	4
Exercise Class or Aerobics	4
Downhill Skiing or Snow Boarding	4
Bowling	2
Baseball or Softball	3
Tennis	4
Weight-training	3
Fishing	3
Volleyball	5
Basketball	6
Soccer	5
Other Activities	4
Other Activities	4
Other Activities	4
Transportation-walking	3
Transportation-bicycling	4

Here we notice that jogging and running are placed at the same category. The MET value for jogging and running is 7 and 12 respectively. The CCHS uses an average value (9.5) of those two to organize the category, so as to be in the line with Campbell's Survey and the Ontario Health Survey.

Table III: Total Daily Energy Expenditure

Variable	Observation	Mean	Min	Max
Total Daily Energy Expenditure	61647	2.3540	0	36.7061

Table 2: Household Income Distribution from All Sources

Household Income	Frequency	Percentage
NO OR <\$20000	6894	11.2%
\$20000-\$39999	13607	22.1%
\$40000-\$59999	11661	18.9%
\$60000-\$79999	8956	14.5
\$80000 OR MORE	20529	33.3%

Source: CCHS 2012 Codebook

Table 3: Physical Activity Distribution

Physical Activity Index	Frequency	Percentage	Cumulative Percentage
Active	17357	28.81	28.81
Moderate Active	15019	24.93	53.75
Inactive	27863	46.25	100.00
Total	60239	100.00	

Source: CCHS 2012 Codebook

Table 4: Concentration Index Decomposed by Various Physical Activities

Name of Activity	Concentration Index	Contribution of Each Activities	Relative Contribution
Daily Average Energy Expenditure	0.0663	n/a	n/a
Walk for Exercise	-0.0271	-0.0078	-11.73%
Gardening or Yard Work	0.0372	0.0046	6.92%
Swimming	0.1069	0.0034	5.08%
Bicycling	0.0612	0.0031	4.70%
Popular or Social Dance	0.0212	0.0003	0.50%
Home Exercise	0.0477	0.0044	6.66%
Ice Hockey	0.3324	0.0068	10.18%
Ice Skating	0.2311	0.0015	2.29%
In-line Skating or Rollerblading	0.0338	0.0001	0.21%
Jogging or Running	0.1780	0.0204	30.72%
Golfing	0.2475	0.0043	6.47%
Exercise Class or Aerobics	0.1437	0.0046	6.99%
Downhill Skiing or Snow Boarding	0.2092	0.0009	1.31%
Bowling	-0.0438	-0.0001	-0.22%
Baseball or Softball	0.1704	0.0009	1.43%
Tennis	0.2161	0.0009	1.35%
Weight-training	0.1925	0.0076	11.44%
Fishing	0.0797	0.0009	1.41%
Volleyball	0.1277	0.0014	2.09%
Basketball	0.0553	0.0011	1.62%
Soccer	0.1389	0.0025	3.75%
Other Activities	0.0886	0.0038	5.74%
Other Activities	0.1504	0.0009	1.32%
Other Activities	0.1855	0.0002	0.26%
Transportation-walking	-0.0348	-0.0011	-1.71%
Transportation-bicycling	0.1585	0.0008	1.22%

Table 5: Concentration Index Decomposed by Demographic Groups

$$CI = CI^W + CI^B + \text{transvariation}$$

1. Male vs Female

Gender	N	CI_{gender}	Mean Value _{gender}
Male	27576	0.0722	2.5821
Female	34071	0.0675	2.1693

Male	Female	Total	CI^W	CI^B	Transvariation
0.0722	0.0675	0.0729	0.0698	0.0049	-0.0018
Relative Contribution			95.74%	6.76%	-2.50%

2. Immigrant - Yes or No

Immigrant	N	$CI_{\text{immigrant}}$	Mean Value _{immigrant}
Yes	8726	0.0357	2.1961
No	51026	0.0768	2.3916

Immi-Yes	Immi-No	Total	CI^W	CI^B	Transvariation
0.0357	0.0768	0.0729	0.0712	0.0019	-0.0038
Relative Contribution			102.74%	2.78%	-5.52%

3. Less than 10 years or more:

Immigrant-year	N	CI_{10}	Mean Value ₁₀
Less than 10	1732	-0.0053	2.5165
More than 10	6,858	0.0617	2.1064

Less than 10	More than 10	Total	CI^W	CI^B	Transvariation
-0.0053	0.0617	0.0310	0.0462	-0.0060	0.0310
Relative Contribution			148.71%	-19.30%	-29.41%

4. Age

Age	N	CI_{age}	Mean Value _{age}
12-17	5104	0.0215	4.3842
18-64	39784	0.0545	2.3899
65+	16759	0.0856	1.6502

12-17	18-64	65+	Total	CI^W	CI^B	Transvariation
0.0215	0.0545	0.0856	0.0729	0.0553	0.0215	-0.0039
Relative Contribution				75.86%	29.49%	-5.35%

Table 6: Achievement Index Decomposed by Activities

$$AI = \mu(1 - CI)$$

Name of Activity	Concentration Index	Mean of Each Activities	Achievement Index
Walk for Exercise	-0.0271	0.6750	0.6933
Gardening or Yard Work	0.0372	0.2904	0.2796
Swimming	0.1069	0.0742	0.0663
Bicycling	0.0612	0.1199	0.1126
Popular or Social Dance	0.0212	0.0370	0.0362
Home Exercise	0.0477	0.2181	0.2077
Ice Hockey	0.3324	0.0478	0.0319
Ice Skating	0.2311	0.0155	0.0119
In-line Skating or Rollerblading	0.0338	0.0097	0.0094
Jogging or Running	0.1780	0.2693	0.2214
Golfing	0.2475	0.0408	0.0307
Exercise Class or Aerobics	0.1437	0.0760	0.0651
Downhill Skiing or Snow Boarding	0.2092	0.0098	0.0077
Bowling	-0.0438	0.0079	0.0082
Baseball or Softball	0.1704	0.0131	0.0109
Tennis	0.2161	0.0098	0.0077
Weight-training	0.1925	0.0928	0.0749
Fishing	0.0797	0.0276	0.0254
Volleyball	0.1277	0.0255	0.0223
Basketball	0.0553	0.0458	0.0433
Soccer	0.1389	0.0421	0.0363
Other Activities	0.0886	0.1011	0.0921
Other Activities	0.1504	0.0137	0.0116
Other Activities	0.1855	0.0022	0.0018
Transportation-walking	-0.0348	0.0769	0.0796
Transportation-bicycling	0.1585	0.0120	0.0101

Table 7: Achievement Index Decomposed by Demographic Groups

$$AI = \mu(1 - CI)$$

1. Male vs Female

Male	Female	Total	Transvariation
2.3956	2.0227	2.1822	-0.0073

2. Immigrant-yes vs no

Yes	No	Total	Transvariation
2.1177	2.2080	2.1993	0.0045

3. Less than 10 years or more

Less than 10 years	More than 10 years	Total	Transvariation
2.5298	1.9764	2.1211	0.0331

4. Age

12-17	18-64	65+	Total	Transvariation
4.2898	2.2597	1.5089	2.1822	-0.0414

7. References

1. Atkinson, A.B., 1970. On the measurement of inequality. *Journal of Economic Theory*. 2, 244–263.
2. Ainsworth, B.E., Haskell, W.L., Leon, A.S. et al., 1993. Compendium of Physical Activities: Classification of energy costs of human physical activities. *Medicine & Science in Sports & Exercise*. 25 (1): 71–80.
3. Beenackers, M.A., Kamphuis, C.B.M., Giskes, K., Brug, J., Kunst, A.E., Burdorf, A., van Lenthe, F.J., King, T., Thornton, L., Bentley, R., 2012. Socioeconomic inequalities in occupational, leisure-time, and transport related physical activity among European adults: a systematic review. *Int J Behav Nutr Phys Act*. 9(1):116.
4. Clarke, P.M., Gerdtham, U.G., Connelly, L.B., 2003. A note on the decomposition of the health concentration index. *Health Economics*. 12:511-516.
5. Craig, C.L., Russell, S.J., Cameron, C., Bauman, A., 2004. Twenty-year trends in physical activity among Canadian adults. *Canadian Journal of Public Health*. 95:59–63.
6. Dowler, E., 2001. Inequalities in diet and physical activity in Europe. *Public Health Nutrition*. 4 (2B), 701-709.
7. Kakwani, N., Wagstaff, A., van Doorslaer, E., 1997. Socio-economic inequalities in health: measurement, computation and statistical inference. *Journal of Econometrics*. 77 (1), 87–104.
8. Le Grand, J., 1985. Inequalities in health: The human capital approach. *Welfare State Program Discussion Paper 1*, London School of Economics.

9. Makdissi, P., Sylla, D., Yazbeck, M., 2013. Decomposing health achievement and socioeconomic health inequalities in presence of multiple categorical information. *Economic Modelling*. 35: 964-968.
10. Makdissi, P., Yazbeck, M., 2014. Measuring socioeconomic health inequalities in presence of multiple categorical information. *Journal of Health Economics*. 34: 84-95.
11. Pan, S.Y., Cameron, C., Desmeules, M., Morrison, H., Craig, C.L., Jiang, X., 2009. Individual, social, environmental, and physical environmental correlates with physical activity among Canadians: a cross-sectional study. *BMC Public Health*. 9:21.
11. Shields, M., 2006. Overweight and obesity among children and youth. *Health Report*. 17(3): 27-42.
12. Seliske, L., Pickett, W., Janssen, I., 2012. Urban sprawl and its relationship with active transportation, physical activity, and obesity in Canadian youth. *Statistics Canada Health Reports*. 23(2).
13. Wagstaff, A., 2012. Inequality aversion, health inequalities and health achievement. *Journal of Health Economics*. 21: 627-641.
14. Wagstaff, A., Paci, P., van Doorslaer, E., 1991. On the measurement of inequalities in health. *Social Science & Medicine*. 33: 545-57.
15. Wagstaff, A., van Doorslaer, E., Watanabe, N., 2003. On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. *Journal of Economics*. 112:207-23.

16. Yitzhaki, S., 1983. On an extension of the Gini index. *International Economic Review*. 24 (3), 617–628.
17. Zhang, Q., Wang, Y., 2004. Socioeconomic inequality of obesity in the United States: do gender, age, and ethnicity matter? *Social Science & Medicine*. 58:1171–80.