Experimental tests of multidimensional inequality measures

What are people' judgments on the distribution of quality of life over income groups?

Richard Folger Zhao – 334477
March 14, 2011

Supervisor: Dr. Tom Van Ourti

Co-evaluators: Dr. K. I. M. Rohde

Dr. Elly Stolk

I Introduction

According to OECD report (2008), over the past two decades, income inequality has widened in more than ¾ of OECD countries. There seems to be a general feeling that inequalities have been on the rise for the recent times both in OECD and non-OECD countries. However, income inequality is only one aspect of inequalities that affects the well-being of the individuals. There are many different dimensions of inequalities such as health, education, or opportunity, which can affect the well being of the individual differently. One of the most important objectives in social policy making is to reduce the inequalities. In order to tackle the issue of inequality, the first step is to identify how to measure inequality. There are many literatures published on how to measure inequality. One of the most common ways to measure inequality is to focus one single aspect of the inequality. Income or health is usually one of the aspects that are mostly talked about in the literature on measurement of inequality. However, well-being consists of multidimensional factors that need to be taken into consideration when measuring inequalities. Concentrating on single dimension when measuring inequality will most likely yield an incomplete picture of the well-being. The most well-known measure of well-being (HDI) the human development index published by the United Nations development programme consists of three dimensions: life expectancy at birth, life attainment, and real GDP per capita.

The literatures on early measurement of multidimensional inequality can be traced back to Kolm's paper using dominance criteria (1977) to assess whether one distribution was more equal than another. Since then, there were numerous papers concentrated on that topic: for instance Atkinson & Bourguignon (1982), Le Breton (1986), Koshevoy (1995, 1998), Koshevoy & Mosler (1996) and Moyes (1999)   The shortingcomings of this dominance analysis is that the number of distributions that can be compared is limited. As a result, it is not the ideal method for policy making because it cannot assess the full effect on the distribution of well-being. (Bleichrodt et al, 2009)

Since dominance criteria only provide partial ordering of the distributions of the dimension, an inequality index was introduced to achieve the complete ordering of all its distributions. Inequality index is a normative approach to inequality measurement in which social evaluation function is used to ranks alternative distribution according to the social preference. Atkinson (1970) and Kolm (1969) were the pioneers for this normative approach to inequality measurement involving single dimensional distribution. Multidimensional inequality measurement was proposed among Kolm (1977), Tsui (1995), and Gajdos and Weymark, (2005).

Although there have been a number of a literatures on the multidimensional inequality measurement, there has not been much discussion on testing the conditions behind the proposed multidimensional inequality indices. It is interesting to see whether or not general public concur that these principles should take precedence of social preferences over multidimensional distribution.

According to Tsui (1999), there are two ways of looking at the multidimensional inequality; first as spread of distribution of the different dimensions and, second, as correlations among different attributes. This paper tests empirically the conditions behind second aspect of the multidimensional inequality. The goals of this paper is to see if the axioms behind correlation between different attributes support the general view of what is considered important by the society. Overall, the data demonstrate there is no strong evidence that subjects in this experiment either do or do not seem to support the conditions behind correlation between different dimensions statistically speaking. Two versions of health conditions were introduced in the questionnaire to test whether there are any consistencies of agreement rates among all subjects that agree with both axioms (Correlation Increasing Majorization and Principle of Income-related health transfer and they were sorted either by income or average duration spent per year in that particular health condition. The result shown that it is inconclusive whether or not most subjects seem to favor the decreasing correlation between the two dimensions in the questionnaire. There also does not seem to have any difference in agreement rates between versions of back pain and versions of depression for both Correlation Increasing Majorization and Principle of Income-related Health Transfer.

In the paper, "Experimental Tests of Multidimensional inequality measures" written by Han Bleichrodt et al (2009), empirical tests were done to assess several of the multidimensional inequality conditions used in literature with two dimensions: Income and life expectancy. In this paper, the concentration of tests is mainly on the correlation between different attributes rather than the dispersion of distribution of the different dimensions. In addition, life expectancy is replaced with quality of life. The main purpose of this study is obtain information on how people value distributions of income and quality of life through empirical tests of multidimensional inequality measure involving correlation between attributes.

The remaining section of this paper is structured as follows: Section two reviews available literatures collected on the topic of different ways to describe quality of life. Section three describes the background theory and conditions that are going to be tested. Section four introduces the questionnaire design process and how the subjects were selected as well as the data collection procedure. Section five discusses the data analysis, result, and also conclusion for the paper. A version of the questionnaire is available at Appendix A.

II Literature Review

This paper tests multidimensional inequality measures involving two attributes: these are income and quality of life. Income is a very self-explanatory concept; there is not a lot of explanation needed for the general subjects to understand monthly income per Euro per person for each distribution. However, quality of life, on the other hand, is rather a complex concept that needs to give some detailed description so every respondent will have a clue what exactly does quality of life entail. The concept of quality of life in this study is based on the quality adjusted life year (QALYs). The reason for choosing QALY concept is because of its wide usage in health care evaluation of health care programs or health technology assessment through using cost per QALY gained to assess the impact of different interventions on both length and quality of life and also quality of life weight is preference-based; more desirable health state receives greater weight than less desirable health state, which is in line with this study to see how respondents prefer one distribution over other in terms of income and quality of life. The basic concept and property of the QALY is explained in the following section. It is relevant for this paper because quality of life indicator used in experiments is mainly based on the QALY properties.

2.1 Quality adjusted life year (QALYs)

Quality adjusted life year (QALY) is a common standard for the outcomes of health care. It enables the assessment of the changes in quality of life (morbidity/Health related quality of life) and quantity of life (mortality) at the same time. It combines morbidity that is the quality of life and mortality, which is the length of life into a single indicator to measure the quality adjusted life year. QALY is equal to quality of life weight multiply by life year. In order to satisfy the QALY concept, the quality of life weight has to be fulfilling three conditions. First, Quality of life weight has to be based on preferences; which means more desirable health state receives greater weight than the less desirable health states. Second, the scale of QALY weight may contains many points, but two points that have to be included are perfect health and death and weights are required for these two points. Last, but not least, intervals of equal length on the scale have equal interpretation.

An article that examines some of the assumption underlying QALY as a whole written by Carr-Hill. (1989) mentioned that QALY enables the decision makers to compare value of different health programs and it reflects social preference. For a patient with a year of full health equals 1 QALY unit and 1 year of current life with lower than full health level as q<1, QALY unit equals q unit. This way the current value of expected remaining QALY could be calculated given the health profile of individual over expected life span. Any health intervention, or treatment that changes the health profile and expected life span gives a different QALY. Any health intervention that yields a higher QALY is considered better. When factoring the costs of each intervention, cost per QALY gained can be calculated for each intervention and intervention with the lowest cost per QALY gained is socially preferable. Now that the concept of QALY is properly explained, the next step is to

explore different indirect utility measurements that express quality of life that can be used for this study. There are two widely adopted indirect utility measurements that are being used in a lot of literatures in health care evaluation.

2.2 Utilization of indirect utility measurement to express quality of life

Indirect utility measurement enables patients to classify their health using a multidimensional descriptive system and a value is assigned to each health state within that system. The underlying assumption is that disease specific problem affects the quality of life on overall domains of health; therefore the value from the generically defined health state can properly express quality of life of the patients in that particular health state. There are numerous validated instruments that can be used to assess the quality of life of a particular health state. As previously mentioned, one of the widely adopted indirect utility measures that have been used in various health care research is EQ-5D.

A. EQ-5D

A group of researchers from the centers in England, Finland, The Netherlands, Norway and Sweden established EuroQol group in 1987 to develop a standardized non-health specific instrument to value and express the health-related quality of life. EQ-5D is a well-know generic instruments allow for measurement of health outcome developed and validated by the EuroQol group. EQ-5D has five dimensions: 1) morbidity, 2) self-care, 3) usual activities, 4) pain/discomfort and 5) anxiety/depression. Each dimension has 3 level of severity: 1) no problem, 2) some problem/moderate problem and 3) severe state. The total possible states under EQ-5D are 245 including unconscious and death and it has utility score ranging from -0.594 to 1. The higher the score, the better the quality of life indication. One of the advantages of using EQ-5D instrument is that it is easy to collect large sample of surveys due to brevity and simplicity. Due to its noticeable advantages such as low administrative burden, cheap to distribute among large sample, many translations, and very sensitive in assessing poor health states, EQ-5D is one of the most frequently used indirect utility instruments. EQ-5D has been used and validated in many health care researches to value quality of life; it ranges from large national studies to specific disease related studies.

In a Norwegian national study led by Erik Nord (1990) to see how health valuation made by Norwegian subjects using EQ-5D instrument compared to the existing studies done by other EuroQol group (UK, Sweden, and Netherlands) for cross-national comparison. The existing EuroQol studies consisted of self-administrated questionnaire and subjects were asked to rate different states and also score their own health and quality of life as well as given their own background information. In initial joint publication the group presented & compared measurement of result from large-scale survey in England, Holland and Sweden. The results showed that valuation of health states across countries were shockingly similar. The Norwegian

study was conducted in Fall of 1989 with 8 random samples of people from 18-84 drawn from central population register. There were some minor differences in this study that questions about one's own health was omitted, and strong pain was used in lieu of extreme pain in other studies and the answers to the survey was not anonymous, in fact, questions about name and address were gathered. In general, Norwegian valuation compared to those of other countries was very similar.

The responsiveness of EQ-5D was also assessed in a research conducted by A.D.I Van Asselt and his group (2008) to compare with Borderline Personality Disorder severity index-IV, a standard clinical measurement for population suffering from Borderline personality disorder (BPD). Data was collected from a multi-center, randomized trail in Netherlands comparing 2 types of outpatient psychotherapy for BPD. The time horizon for the experiment was total of 4 years with assessment every 3 month during the first 3 years and final follow-up assessment after the 4th year. The standardized response mean is higher for BPDSI for all instruments and it is largest effect is in the 1st year. Incremental effect for 2nd year is smaller. For both EQ-5D & BPDSI, the standardized response mean after 3rd year is smaller than after 2nd year. EQ-5D utility score discriminated significantly between "responders" and "non-responders" to the therapy in trial. The study concludes that EQ-5D can be used as a useful tool for assessing quality of life in-patient with BPD. EQ-5D seems to be the most appropriate instrument for describing quality of life in this study because of its extensive validation in numerous published literatures.

2.3 Elicitation of different health states in measurement of health utility

The concept of QALY is not like income, which is self-explanatory; it is complex for the general public who will be filling out questionnaire to comprehend. In order to describe the QALY concept easy-to-understand for everyone who will be filling out this questionnaire, few articles written by Han Belichrodt was review to see how the author describe different health states to the general public using EQ-5D instrument.

A study by Abellan-Perpinan et al (2009) pertaining to the predictive validity of prospect theory versus expected utility in health utility measurement, two health states from utility of EQ-5D were labeled A and B. The description of both health states is straightforward using 5 dimensions and 3 levels within each dimension from EQ-5D instrument. (See table below) The respondent in the study were presented a series of questions pertaining to both health states and asked to make choice of preference between the two health states. It is clear that state A dominates

The description of health states A and B.

Health state A	Health state B
Some problems walking about	Some problems walking about
Some problems performing self-care activities (e.g. eating, washing, dressing)	Some problems performing self-care activities (e.g. eating, washing, dressing)
No problems performing usual activities (e.g. work, study, family or leisure activities)	Unable to perform usual activities (e.g. work, study, family or leisure activities)
Moderate pain or discomfort	Moderate pain or discomfort
Moderately anxious or depressed	Moderately anxious or depressed

state B in the sense that it yields a level of functioning that on each dimension is at least as good as the corresponding level B. The main goal of this paper is to test whether prospect theory leads to better health evaluation than expected utilities when risk are introduced.

In the article "New tests of QALYs when health varies over time" by Bleichrodt et al (2008), the authors selected 4 health states from 4 EuroQol health stares and described them in a rather colorful way. Each health state was assigned a color to aim to facilitate decision-making by reminding the subjects of relative attractiveness of the health states. The 4 health states were labeled K, L, M, and N and their corresponding colors were green, yellow, orange, and red. Each health state has 5 dimensions and within each dimension there are three levels of severity according to the EuroQol. Subjects are asked to rank from the best state to the worst health state. Please see the figure below for the 4 different health states.

Health state K

In a health state K, your health is characterized by:

Mobility	**No problems** walking about
Self-Care	**No problems** with self-care
Usual Activities	**No problems** with performing usual activities (for example, work, study, housework, family, leisure activities)
Pain/Discomfort	**No pain** or **discomfort**
Anxiety/Depression	**Not anxious** or **depressed**

Health state L

In a health state L, your health is characterized by:

Mobility	**No problems** walking about
Self-Care	**No problems** with self-care
Usual Activities	**No problems** with performing usual activities (for example, work, study, housework, family, leisure activities)
Pain/Discomfort	**Moderate pain** or **discomfort**
Anxiety/Depression	**Not anxious** or **depressed**

Health state M

In a health state M, your health is characterized by:

Mobility	**No problems** walking about
Self-Care	**No problems** with self-care
Usual Activities	**No problems** with performing usual activities (for example, work, study, housework, family, leisure activities)
Pain/Discomfort	**Moderate pain** or **discomfort**
Anxiety/Depression	**Moderately anxious** or depressed

Health state N

In a health state N, your health is characterized by:

Mobility	**Some problems** walking about
Self-Care	**No problems** with self-care
Usual Activities	**Some problems** with performing usual activities (for example, work, study, housework, family, leisure activities)
Pain/Discomfort	**Moderate pain** or **discomfort**
Anxiety/Depression	**Moderately anxious** or depressed

The main purpose of this article was to test on QALY model when health varies over time. The result support the use of QALY on the aggregate level (economic evaluation of health care programs), but individual level there seems to be less support for QALYs.

In both articles "A consistency test of time trade-off" by Bleichrodt et al (2003) and "Resolving inconsistencies in the utility measurement under risk: tests of generalization of expected utility" by Bleichrodt et al (2005), the authors used similar approaches as the previous articles in which two health states were selected and labeled A and B from the EuroQol EQ-5D instrument along with the 5 domain descriptions to describe the health state of individuals for the subjects recruited from the experiments. Subjects were then asked a series of questions related to either health states and to make trade off between two health states.

After reviewing numerous articles on QALY instruments, it is obvious that using EQ-5D is most suitable for this experiment due to its brevity, easy-to-understand description and wide validity in many different health conditions as previous mentioned. However, there are some changes made on this paper pertaining the utilization of EQ-5D instrument. In this paper, rather than listing different five dimensions separately, it is combined into a paragraph with more detailed description expanded in each dimension, which paints a bigger picture for the subjects regarding particular health condition. In this paper, two specific health conditions were selected and there were severe back pain and common depression.

2.4 Multidimensional inequality measure through correlation among different dimensions.

As mentioned earlier, the most common inequality in the past has been primarily concerned with one dimension indicator, which usually involves income or health. Income is not the best proxy for measurement of well-being, in fact, one dimension indicator whether it's income or health cannot correctly assess the overall well-being of the individual in measurement of inequality.

Kolm (1977) was the pioneer in measurement of earlier multidimensional inequality. His paper is mainly concerned with developing dominance criteria for ranking multivariate distributions. When there are multiple dimensions of well-being being compared, one distribution maybe more equal than the other one if former one shows less dispersion than the latter one and the mean of two distribution are kept the same. However, dominance criteria provide only partial ordering of the possible distributions of different attributes, therefore, an inequality index is more socially preferable because it enables the complete order of all distributions. There are two ways to look at the multidimensional inequality: dispersion in the distributions of different dimensions and correlation among different dimensions. (Tsui, 1999) In this paper, the focus is on measurement of multidimensional inequality through looking at correlation among different dimensions. The two axioms that will be tested in this paper are Correlation increasing majorization proposed by Boland and Proschan (1988) and principle of income-related health transfer. Both axioms are dealing with correlation among the different dimensions and based on the theory, it is socially preferable to have decrease correlation between dimensions for multidimensional inequality

measurement as long as both attributes are desirable and related to each other such as health, income, education attainment or job opportunity. In the previous study done by Bleichrodt et al (2009), the authors tested several assumptions underlying both types of the multidimensional inequality measures to see which one mostly connects with general public's preferences. The result showed that people mostly care about the decrease in the dispersion of dimensions rather than the correlation among different dimensions. In this paper there are few modifications had been made in regards to the health component of dimension. Instead of using the general life expectancy as the other dimension, quality of life component was used in lieu of life years. This paper also mainly focuses on the correlation between the two dimensions that are income and quality of life. In the next section, background information, two axioms are being introduced and explained: there are Correlation increasing majorization and Principle of income-related health transfer.

III Background information

3.1 Correlation increasing majorization

Correlation increasing majorization is used to assess the correlation among the different dimensions of the well-being.

Correlation increasing majorization concept was proposed by Boland and Proschan (1988) and it states that for all the allocation (x,y) and (x',y'), if (x',y') is obtained from (x,y) through a series of correlation increasing majorization then (x,y) is more socially preferable than (x',y') and the allocation (x,y) has more equitable distribution attributes.

Referring back to the previous allocations, (x,y) and (x',y'). The dimensions in this case are quality of life represented by x and income represented by y. Based on the principle of Correlation Increasing Majorization, the rearrangement of the allocation for both individuals are done in a such fashion that one of the individuals will have at least as equitable numbers of quality of life weight and amount of income as the other and strictly more of one of these and the rearrangement is not a permutation of the individuals. If person A's quality of life weight is greater then person B's quality of life weight for two individuals A, and B, then the rearrangement income will be done in a way such that y' of Person A is Maximum of (Person A income, Person B income) and y' of Person B is Minimum of (Person A income, Person B income) or if the person A income is greater than Person B income for both individuals A and B, then the rearrangement will be done to quality of life x' of Person A = Maximum of (Person A QALY, Person B QALY) and x' of Person B =Minimum of (Person A QALY, Person B QALY)

3.2 Principle of Income-related health transfer

The principle of Income-related health transfer states that a transfer of health from one individual who is in better income to another who is in lesser income does not lead to a decrease in social welfare as long as the transfer does not change the income ranking of the individuals.

The principle of income-related health transfer suggests correlation increasing majorization as transfer of health from wealthier person to poorer individuals reduce the relationship between the income and health.

Income-related health transfer requires the transfer of health from someone who is better-off in terms of income to someone who is worse-off in terms of income does not lead to reduction in social welfare given that the transfer does not change the initial ranking of the individuals in terms of income ranking. This income-related health transfer raises question because what if the individual who is better-off in terms of income is in fact in worse-off health state then the individual who is worse-off in terms of income is actually in better-health states? The income-related health transfer in this case does not seem socially preferable. In order for principle of income-related health transfer to be more acceptable, it is important there are strong relationship between the health and other attributes such as income, because

one cannot assume that individual who is richer will indeed be healthier or vice versa.(Bleichrodt & Van Doorslaer, 2006)

It is more desirable to study the multivariate concept of the inequality because focusing on one dimension does not give an accurate picture of inequality measurement, but yet well-being consists of multidimensional aspect. Therefore, the policy maker cares not only the distribution of the health, but also the distribution of other attributes such as education attainment, income etc... to give a big picture of overall inequality.

IV. Questionnaire design

The questionnaire design started off with a brief explanation of the purpose of the survey that is to obtain information on how people value distributions of income and average number of days in a year spent in certain health conditions. There were two health conditions, severe back pain and depression, which were selected based on the nature of the condition. This is to assess whether there was a difference between how people value the physical health condition and the mental health condition. The detailed description of severe back pain and common depression is depicted as the following:

A) Severe back pain
People with severe back pain have no problem walking and moving around on their own; there is no assistance necessary for mobility. They neither need assistance to wash, dress or take care of themselves. However, severe back pain affects performance in usual activities: e.g. they experience some problems in sitting down for a long period of time or in studying without taking some breaks in between. Also they have some problem doing housework such as cleaning and doing dishes for a long period of time. In addition, people experiencing severe back pain have a moderate degree of pain or discomfort: e.g. the pain can be a dull ache, a sharp or piercing or burning feeling and sometimes it could have other symptoms such as weakness, numbness or tingling feeling. There is no psychological effect associated with individuals with severe back pain; they are not anxious or depressed.

B) Depression
Depression is a health condition that affects mostly the mind and reasoning. Most people experience sadness, loneliness, loss of interests, feeling of tiresome, headache and anxiety at one point in their lives, and it could be considered as normal reaction to loss, life struggles, or other unfortunate incidents. The common symptoms for depression are difficulty in concentrating, decreased energy, feeling of worthlessness, difficulty in sleeping, possible weight loss, and headaches. People who have depression have no problem in walking and moving around on their own; there is no assistance required for mobility. They also don't require any assistance with washing, dressing or taking care of themselves. However, people with depression might have some problems in performing usual activities. They might encounter problems when it comes to studying or working because of their difficulties in staying focused, remembering details, or decision-making. Depression could also trigger loss of interest or pleasure in hobbies, and activities once enjoyed. Moreover, people suffering from depression usually have a moderate degree of pain or discomfort such as headaches, cramps, or digestive problems that don't go away easily. Last but not least, people with depression generally are moderately anxious or depressed, which includes feelings of restlessness, emptiness, or being annoyed.

The reason that there were two versions of each health condition was to avoid heuristics; it is interesting to see whether sorting of attributes make a difference in terms of agreement. Within each version of health condition it was further sorted

either by income or average number of days spent in that particular health condition to statistically examine whether heuristics make a difference. The main goal of this experiment is to test multidimensional inequality measures and to see what are people's judgments on the distribution of quality of life over income groups.

In each version of the questionnaire, subjects were asked to imagine a country called Alfaland for which two governmental policies have been proposed that have identical impact on the country, except for the income and average number of days in a year spent in either back pain or depression depending on the version of the questionnaire. Both income and quality of life were selected because they are important well-being dimensions that are easy to understand for most of the subjects that were selected. Average number of days in a year spent in particular condition is the quality of life indicator that has two components: It combines type of illness and average duration spent in illness and make them into one indicator to measure quality of life.

Subjects were then informed that the population of Alfaland consists of 5 equally sized groups of people and all groups have constant life expectancy and same average age. The only difference among the groups is the income and the average duration of the particular health condition. Within every group the income and average number of days spent in that particular condition are identical between different members of the same group. An example of the table was given in each of the questionnaire with brief explanation of 5 different groups as well as income and average duration of health condition.

In the following table, the five numbered columns represent the five groups of the people in that country.

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	13	8	5

The five columns, named 1, 2, 3, 4, and 5, give the income and average number of days in a year spent in severe back pain for each group. Income is the net monthly income in Euros per person. The price level in Alfaland will be similar to the price level in the Netherlands in 2010. The average number of days in severe back pain does not affect the corresponding income. Underneath, we provide a description of severe back pain.

The income corresponding to each group is the net monthly income in Euros per person. The baseline distribution of monthly income in Euro per person was based on   the (2001) Dutch wave of the European Community Household Panel and adjusted for inflation. The price level in Alfaland will be similar to the price level in the Netherlands in 2010. All groups experience some sort of health condition (either back pain or depression) at one point in their lives which was described as the average number of days per year in particular health condition. On all days without severe back pain or depression, the individuals resume to full health.

Depending on the version of the questionnaire, either back pain or depression was one of the health conditions being described in a very detailed paragraph with 5 dimensions of EQ-5D using a very easy-to-understand language. Subjects were asked to repeatedly choose between two distributions of income and average number of days per year in particular health condition which were labeled under Policy A and Policy B.

The reason for naming this country Alfaland was to give the subjects an unbiased view on the whole distribution and making them to take on the perspective of an impartial decision-maker.

How are different versions of the questionnaire presented?

There are a total of 20 questions in each version of the questionnaires; 10 questions are to test the principle of Correlation increasing majorization and the other 10 questions are to test the principle of Income-related health transfers. Questions were arranged in a way that it started off with CIM question and then it followed with PIRHT and so on. At each question, the policy that agrees with axiom being tested was also rearranged randomly so there were no specific patterns.

All versions have the same arrangements in terms of the way the principle of CIM and PIRHT were presented in the question format as well as the policies that corresponded with the axiom being tested. Please see the table below for the order that each principle was presented as well as the policy that agrees with the principle.

Question	Nr/Theories tested	Policy
1	1. CIM	B
2	10. PIRHT	A
3	2. CIM	A
4	9. PIRHT	B
5	3. CIM	A
6	8. PIRHT	B
7	4. CIM	B
8	7. PIRHT	A
9	5. CIM	A
10	6. PIRHT	B

11	6. CIM	B
12	5. PIRHT	A
13	7. CIM	B
14	4. PIRHT	A
15	8. CIM	B
16	3. PIRHT	A
17	9. CIM	A
18	2. PIRHT	B
19	10. CIM	A
20	1. PIRHT	B

As previously mentioned, there are two versions of Back pain as well as two versions of depression. Within the versions of Back pain and Depression, the only difference was the sorting of two attributes; Back pain was sorted by income first versus Back pain was sorted by average number of days per year in particular health condition. The same goes for the versions of depression to assess if there was any difference in the responses when income and quality of life had been switched in the order it was presented.

V. Data collection procedure

As previously mentioned, there were four versions of the questionnaire and there were Back pain sorted by income, Back pain sorted by health (average number of days per year spent in particular health condition), depression sorted by income and depression sorted by health.

An email roaster of some of Health Economics, Policy and Law (HEPL) Master students from Institute of Health Policy and Management of the Erasmus University of Rotterdam in the school year of 2009-2010 as well as the current year (2010-2011) students from the same program were obtained from Blackboard. A total of 190 HEPL students from Erasmus University of Rotterdam was chosen and randomly assigned a version of the questionnaire. The method of the distribution was through email communications.

Another email roaster of 28 graduates of Bachelors degree from the City University of New York (CUNY) was chosen and randomly assigned a version of the questionnaire and the method of distribution was a mix of email communication and also filing in the paper versions of the questionnaire.

The goal is to achieve a minimum of 60 surveys in total in which 15 surveys are required per version. There were 37 completed surveys from the HEPL program and 23 completed surveys from City University of New York (CUNY). On average it took around 20-30 minutes for subjects to complete the survey and the survey was written in English.

VI. Data Analysis

A total of 60 completed questionnaires were collected. Of the 60 surveys, there were 4 different versions and each version had 15 surveys. Answers were extracted from the surveys and inputted into a table which contains the following columns of information: Assigned number, Gender, Name of respondent, Date received, Number of questions in the survey, and version of the survey.

Once all the data were inputted into the table, the shares of subjects that agree with each of the axiom being tested were calculated for each question. For clarity, a separate table and scattered plots were also created with percentage of subjects agreeing with each principles being test per question to see overall whether or not the majority of the subjects agreeing with both Correlation Increasing Majorization and Principle of income-related health transfer principles. Each question was also analyzed with T test to see whether the proportion of the subjects behaving in line with each condition differed significantly from 50% from at least 5% significance level. The reason for testing each condition to see if there are at least 50% different is because it is the minimum standard that could be expected if subjects were indifferent or their choices were purely random. If the agreement rate reaches 50% for each condition being tested then it is sensible to make any reasonable conclusion. The percentages of subject that agree with each if the axiom being tested were also calculated by version.

In order to assess whether different health condition could contribute to differences in agreeing with each of the principles being tested, a table and column graph were created for each principle with percentages of agreement from the subjects for both health conditions. A two-sample T test was done at minimum 5 % significance level to test each question of back pain and depression to see if the differences between the agreement rates are statistically significant.

To look further into the percentages of agreement of each axiom being tested to see if ranking of different attributes has made a difference, comparison was done within each health condition between version sorted by income and version sorted by average number of days per year in back pain or depression. In order to conclude whether the differences between the agreement rates are statistically different from each other, a two-sample T-test again was done to each question within back pain and depression between version sorted by income and version sorted by average number of days. As mentioned in the background information that Principle of income-related health transfer is similar to Correlation Increasing Majorization that both principles support reducing the inequality through decreasing correlation between the different dimensions of well-being. Hence there are CIM questions that are very similar to PIRHT questions and it is interesting to look at those questions and see if the agreement rates are somewhat consistent. The similarity of the two conditions was determined based on the income groups that are in comparison as well average duration of health condition either before or after the PIRHT health transfer. A-two sample T-test was done to similar CIM and PIRHT questions to

examine whether the agreement rates are significantly different from each other at a minimum 5% level.

Odd numbered-questions were to test the Correlation Increasing Majorization principles and even numbered-questions were to test the Principle of income-related health transfer. For simplicity of the analysis, each principle was sorted by the test number rather than the actual question number.

VII. Result

The majority of the respondents were female (45) compared to 15 male respondents. Most of the completed surveys were collected in the month of November (31), followed by the month of January which yielded 25 completed surveys and in the month of December only 4 surveys were collected.

As the share of the subjects that agreeing with CIM and PIRHT were calculated for each questions, the result shows it is inconclusive whether or not that majority of the respondents are concerned with decreasing correlation between the dimensions.

Table 1 and figure 1 show the result for percentages of subject that agrees with Correlation Increasing Majorization for all the tests that concern with the axiom. It seems that overall, majority of the respondent support the principle of CIM. The proportions of subjects consistent with CIM vary between 42% and 68%. Test 2 and Test 5 seem to have the strongest support, which yield 68% and 65% agreement rate respectively. Test 1, 3, 6, 7, and 9 fluctuates slightly around 50% agreement rate for CIM. Test 8 scored around 42%, which is considered the lowest agreement rate. However, after performing one sample T test on different CIM test to see how many CIM tests are actually different from 50% agreement rate at a minimum (5%) significance rate, the results are shown on table 2. It seems that only tests 2 and test 5 are statistically significant; test 2 is significantly different from 50% at P value=0.0036, and test 5 scores a significance at P=0.0188. Since there are only two tests that shows proportion of subject agreeing with Correlation Increasing Majorization to be significant not likely due to chance, it is inconclusive in rest of the tests whether the strong agreement rates or disagreement rates may because of choices are purely random. Moreover, it seems that tests of CIM that yields the two highest agreement rates involves smaller difference in average numbers of days per year between the two groups in comparison. It can be concluded that when it comes to smaller difference in average duration of health condition between the groups in comparison, the subjects are significantly concurring with the correlation increasing majorization. Table 1 also examines all the CIM questions to see which tests show proportion for subjects satisfying CIM if the choice were not purely random (statistically significance). According to table 1, only test 2 and test 5 demonstrates it is unlikely for the proportion of subjects satisfying the CIM questions to have occurred purely by chance. It would be interesting to see later on if the sorting of the attribute plays a role in this difference.

Table 1: Proportion of subjects agrees with CIM and P-value for one-sample T-test of difference from 50%

| | | CIM | |
NR	Test	Proportions	P-50% T-test
1	1	53%	0.609715217
3	2	68%	0.003655206
5	3	52%	0.798702558
7	4	55%	0.443205056
9	5	65%	0.018824491
11	6	47%	0.609715217
13	7	47%	0.609715217
15	8	42%	0.199219153
17	9	50%	1
19	10	55%	0.443205056

Proportion: Percentage of subjects agreeing with CIM
P-50% T-test: P-value for one sample t-test of difference from 50% for CIM

Figure 1: Proportion of subject agreeing with CIM

Table 2 and figure 2 show the percentage of respondents behaving in line with Principle of income-related health transfer per test. Overall, the percentage of subjects agree with PIRHT are somewhat consistent across the board (fluctuates around 55%) with the exception of test 9 and 10 which yield 35% and 38% respectively. After performing a single sample T test to see how different each questions are from 50% at a significance rate of 5 % (table 2), it seems that only tests 6 and test 9 are different at 50% agreement rate at a statistically significant at

a 5 % level. Test 6 has 65% agreement rate and based on t-test it is not likely the choices are due to chance, and test 9 has an agreement rate only 35%, which is below 50% minimum standard for satisfying PIRHT and based on t-test, it can be concluded the result is not likely due to random choice. All other test, the proportions of agreement rates are not significantly different from 50% at 5% significance level. Test 6 has the highest agreement rate for PIRHT that is not likely due to chance and Test 9 has the lowest agreement rate (disagreement) for PIRHT that is not likely due to chance.

Table 2: Proportion of subjects agrees with PIRHT & P-value for one sample t-test of difference from 50%

| | | PIRHT | |
NR	Test	Proportion	P-50% T-test
2	1	62%	0.070332883
4	2	58%	0.199219153
6	3	57%	0.30564365
8	4	60%	0.122250936
10	5	55%	0.443205056
12	6	65%	0.018824491
14	7	53%	0.609715217
16	8	57%	0.30564365
18	9	35%	0.018824491
20	10	38%	0.070332883

Proportion: Percentage of subjects agreeing with PIRHT
P-50% T-test: P-value for one sample t-test of difference from 50% for PIRHT
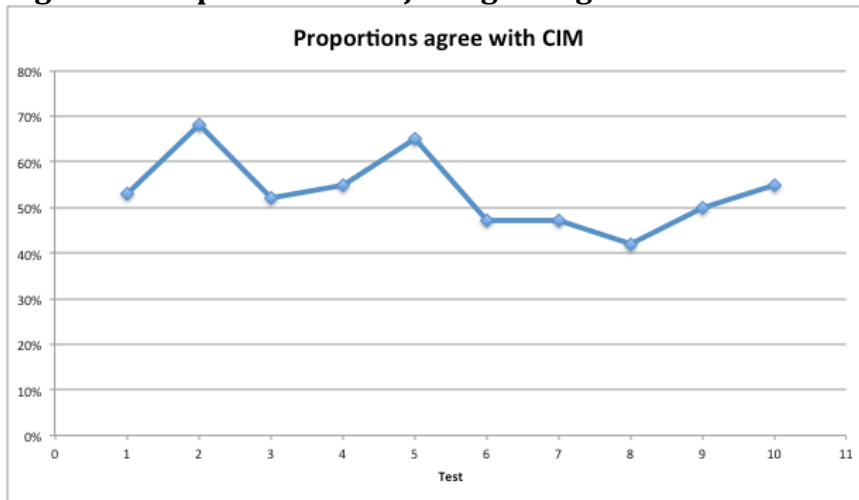
Figure 2: Proportion of subjects agree with PIRHT
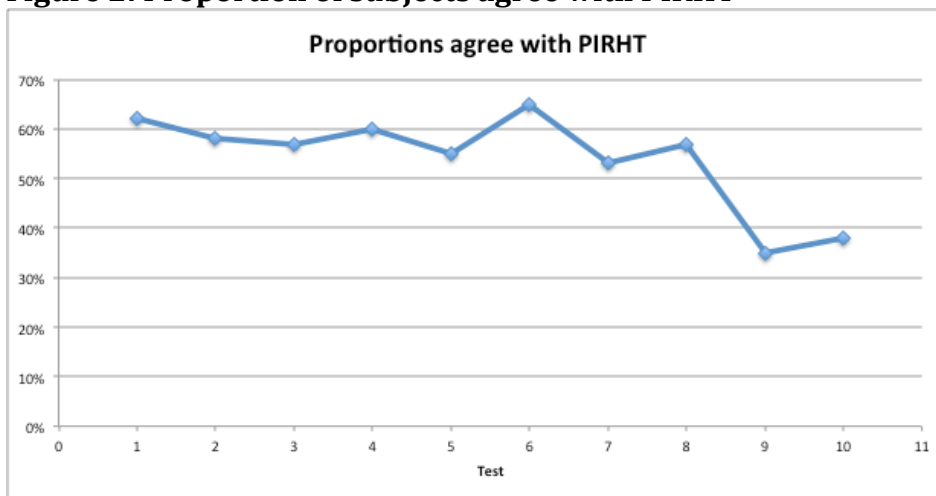
Compare the share of subjects agree with axiom being tested between health conditions (Back pain Vs. depression)

The aim of comparing the proportion of subjects that agrees with CIM and PIRHT between back pain and depression is to see whether the different type of health condition makes a difference in agreement rate. It is obvious that back pain is mainly dealing with physical condition as oppose to depression that is mainly affecting the minds. Two-sample independent T-test is done to every Back pain and depression questions to see if there is any statistical significance between the two types of test that could affect CIM and PIRHT.

When it comes to principle of Correlation increasing majorization, there are no large variation between the proportions of subjects in agreement in back pain version and depression version. On average, it varies around 4-13% between the 2 health conditions with the exception of test 3 & 4, in which the differences are 23% & 24% respectively. (See figure 3 and table 3) After performing T test of CIM between back pain and depression, (table 4), there does not seem to be any different in terms of agreement rates from each other with the exception of test 3 which yields P value=0.051 for T test. Hence, the confidence interval between back pain and depression for test 3 slightly overlaps by small decimal places on the left side. (BP=0.40+/-0.1783 (0.2217-0.5783), DP=0.63+/-0.1715 (0.4547-0.8053)). There is no conclusion that can be drawn as why test 3 yields such a statistic difference between back pain and depression version of the questionnaire.

Table 3: % agreeing with CIM between Back pain and depression

NR	Test	CIM % per Back pair	% per Dep	% differenc
1	1	50%	57%	7
3	2	63%	73%	10
5	3	40%	63%	23
7	4	43%	67%	24
9	5	60%	70%	10
11	6	40%	53%	13
13	7	50%	43%	7
15	8	47%	37%	10
17	9	57%	43%	14
19	10	57%	53%	4

Figure 3: proportion of agreeing with CIM Between Back pain and depression
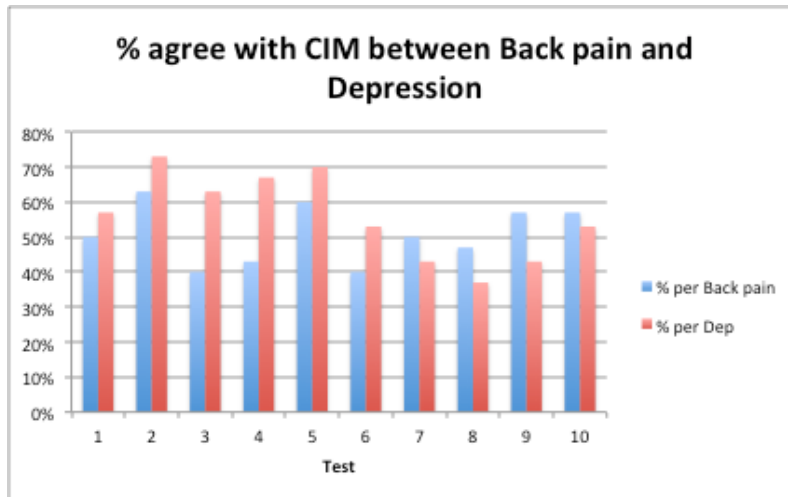
% agree with CIM between Back pain and Depression

Table 4: T test and Confidence Interval between BP and DP concerning CIM

T test and confidence interval between Back pain & Depression

NR	Test	T-test BP Vs Dep.	Interval BP	Intercal DP
1	1	0.611994012	0.50+/-0.181	0.43+/-0.1803
3	2	0.413700569	0.63+/-0.175	0.73+/-0.1609
5	3	0.050726109	0.40+/-0.178	0.63+/-0.1715
7	4	0.071318251	0.57+/-0.180	0.33+/-0.1715
9	5	0.425397898	0.60+/-0.178	0.70+/-0.1668
11	6	0.308727738	0.60+/-0.178	0.47+/-0.1815
13	7	0.611994012	0.50+/-0.182	0.57+/-0.1804
15	8	0.44068105	0.53+/-0.181	0.63+/-0.1754
17	9	0.309813614	0.57+/-0.180	0.43+/-0.1804
19	10	0.799408986	0.57+/-0.180	0.53+/-0.1816

T-test BP Vs Dep: P-value for T test between sample of back pain version and sample of depression version

Interval BP: Confidence Interval for Back pain using Mean +/- confidence value

Interval DP: Confidence Interval for Depression using Mean +/- confidence value

Figure 4 and table 5 clearly show that when it comes to principle of income-related health transfer, the 2 health conditions don't seem to make a difference in terms of agreement rate of the axiom being tested. The percentages vary around 0-16%. After performing two –sample T-test for all the tests between back pain and depression, it shows that overall; there is no significant difference in the agreement rate between the two versions (table 6). Subjects were consistent in valuation when it comes to PIRHT between back pain and depression. It can be states there is no significant difference in agreement rate of PIRHT when it comes to versions of back pain and versions of depression.

Table 5: Proportions agreeing with PIRHT between back pain and depression

NR	Test	PIRHT % per Back Pain	% per DEP	% difference
2	1	60%	63%	3
4	2	57%	60%	3
6	3	50%	63%	13
8	4	57%	63%	6
10	5	50%	60%	10
12	6	57%	73%	16
14	7	53%	53%	0
16	8	53%	60%	7
18	9	37%	33%	4
20	10	40%	37%	3

Figure 4: proportion of agreeing with PIRHT between back pain and depression

% agree with PIRHT between Back pain and Depression
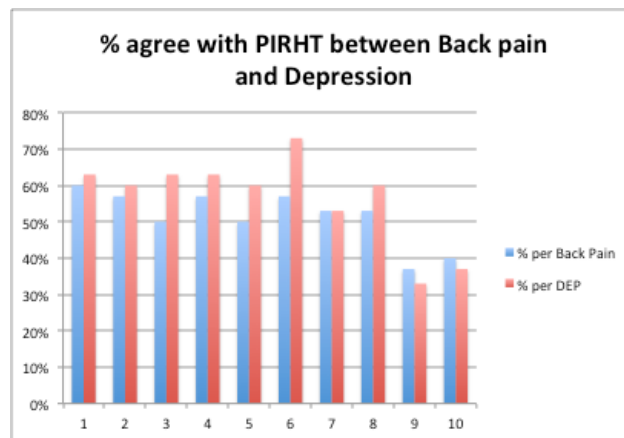
Table 6: T-test and confidence interval between BP & DP concerning PIRHT

T-Test & Confidence Interval Back pain Vs Depression

NR	Test	T-test BP Vs I	Interv BP	Interval dp
2	1	0.79484932	0.6+/-0.1783	0.63+/-0.1754
4	2	0.79762186	0.43+/-0.180	0.40+/-0.1783
6	3	0.30544125	0.50+/-0.1820	0.37+/-0.1754
8	4	0.60546584	0.57+/-0.0.180	0.63+/-0.1754
10	5	0.44483277	0.50+/-0.1820	0.40+/-0.1754
12	6	0.18189221	0.57+/-0.1804	0.73+/-0.1609
14	7	1	0.53+/-0.1816	0.53+/-0.1816
16	8	0.60958397	0.53+/-0.1816	0.60+/-0.1783
18	9	0.79096896	0.63+/-0.1754	0.67+/-0.1716
20	10	0.79484932	0.60+/-0.1783	0.63+/-0.1754

T-test BP Vs Dep: P-value for T test between sample of back pain version and sample of depression version
Interval BP: Confidence Interval for Back pain using Mean +/- confidence value
Interval DP: Confidence Interval for Depression using Mean +/- confidence value

After conducting sensitivity analysis between back pain and depression on two axioms being tested, it can be concluded that different type of health condition does not in general make a big difference in terms of agreement rate between the two principle. Now, it is interesting to see if the ranking of each dimension makes it a difference statistically.

Does ranking make a difference between the agreement rates of CIM and PIRHT in the same health condition?

Within each health condition, two versions of the questionnaire were created based on sorting by income and average duration of health. It is interesting to see whether sorting has made a difference in the proportion of subjects agreeing with both principles. Each principle was examined separately with different versions of back pain and depression.

Figure 5 illustrates the comparison of the proportion of subjects agreeing with CIM between two versions of the back pain questionnaires. Overall, it seems that back pain sorted by average duration of health version yields a stronger agreement rate with CIM compares to back income with the exception test 4 and test 9. However, in order to see statistically how different back pain sorted by income versus back pain sorted by average duration in each of the tests, a two sample independent T test is calculated using two tails and assuming unequal variance. A two-tailed t-test used to reject null hypothesis when the value of test statistic is either significantly small or large. Table 7 illustrates the P-values from the two-sample independent t-test between Back pain version sorted by income and back pain version sorted by average duration of health. Usually, p-value larger than 5% significance level will not reject the null hypothesis that their means are not different from each other. However, the observation in this experiment is limited, therefore it is reasonable to accept significant level at 10%. Based on table 12, within the back pain condition, there is no significant difference in agreement rate between CIM tests ranked by income and CIM test ranked by average number of days spent in back pain with the exception of test 7 which is considered significant at 10% level. According to the two-tailed T-test assessing significant difference at 10% level. P-value is, for the most of the part, presented for 2 tails test because it is aimed to see there is any difference in terms of agreement rate between the two samples which are back pain income and back pain health.

Figure 5: proportion of subjects agreeing with CIM between back pains sorted by income vs. back pain sorted by duration of the health
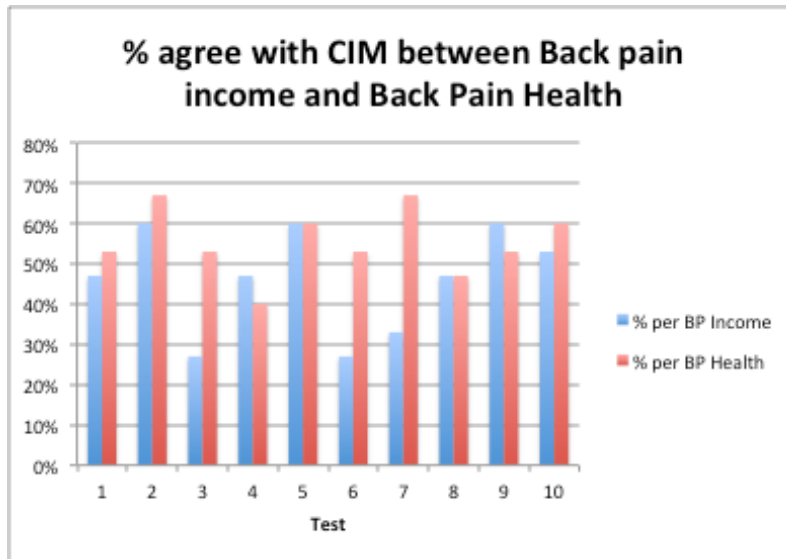
Table 7: P-values from T test between BP income and BP health with 2 tails

T test BP inc vs BP health

Test #	P-BP inc Vs Bp hea
1	0.726321874
2	0.716456215
3	0.145829144
4	0.723952141
5	1
6	0.145829144
7	0.071854237
8	1
9	0.723952141
10	0.723952141

P-BP inc Vs Bp Hea: P-value of 2 sample independent t-test between the sample of Back pain income and sample of back pain health.

According to figure 6, CIM test between depression ranked by income versus back pain ranked by average duration spent in that condition, it seems that depression sorted by income yields a consistent higher percentage per test comparing to the depression sorted by health. However, after computing the P-value for the two-sample independent T test with 2 tails and assumption of unequal variance, it shows (table 8) only test 1 and test 7 have shown significant different (10% level) in terms of agreement rate of back pain sorted by income versus back pain sorted average number of days when it comes questions of Correlation Increasing Majorization. In those cases, proportion of subjects agreeing with CIM of depression ranked by income is considered significantly higher than proportion of subjects agreeing with CIM of depression ranked by average duration of health. All other tests, it shows that there is no major difference in agreement rates between version of depression

questionnaire sorted by income and version of depression questionnaire sorted by health.

Figure 6: Proportion of subjects agreeing with CIM between depressions ranked by income vs. depression ranked by health
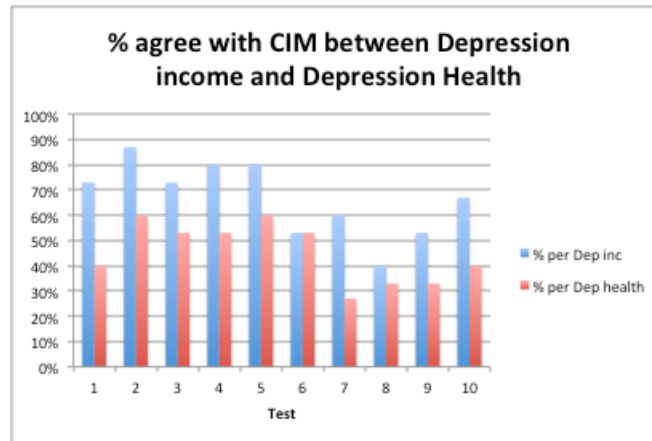
Table 8: T test between DP Inc vs. DP healths two tails

T Test DP inc vs DP health

Test #	P-DP inc vs DP healt
1	0.069284812
2	0.106761049
3	0.271320576
4	0.130430335
5	0.247066653
6	1
7	0.069284812
8	0.716456215
9	0.284919761
10	0.153373684

P-DP inc Vs DP Hea: P-value of 2 sample independent t-test between the sample of depression income and sample of depression health.

When looking at figure 7 and figure 8, there seem to be some differences in agreement rate between sorting of health condition by income or average number of days for Principle of Income-related health transfer. Although, from looking at both figures, it seems that when sorting by income, the proportions of agreement rates

are slightly higher in most of the case. However, when T tests are computed for both conditions in terms of sorting by income versus sorting by average numbers of days, it turns out that overall, there is no significant difference in agreement rates between back pain sorted by income or average number of days with the exception of test 6, which yields p value=0.069 satisfy 10% significance level. All other tests between back pain version sorted by income and back pain version sorted by average duration of health does not give any reason to believe that there are difference in agreement rates for PIRHT. When it comes to (table 9) depression versions of the questionnaire, test 1, 4, and 7 show the possibility that there are some differences between sorting by income versus sorting by average number of days when it comes to depression. Among these tests, depression sorted by income again shows statistical higher agreement rates when compared to agreement rates sorted by average duration of health. It is peculiar that the same test does not seem to come up significant in the versions of back pain. However, at this time, there is no valid conclusion that can be drawn about this difference.

Overall all, across both health conditions, it can be stated that it is not conclusive for both CIM and PIRHT whether sorting plays a big role in the differences between the share of the subject that agrees with the axioms being tested.

Figure 7: % PIRHT between BP incomes vs. BP health. Figure 8: % PIRHT between DEP incomes vs. Dep health
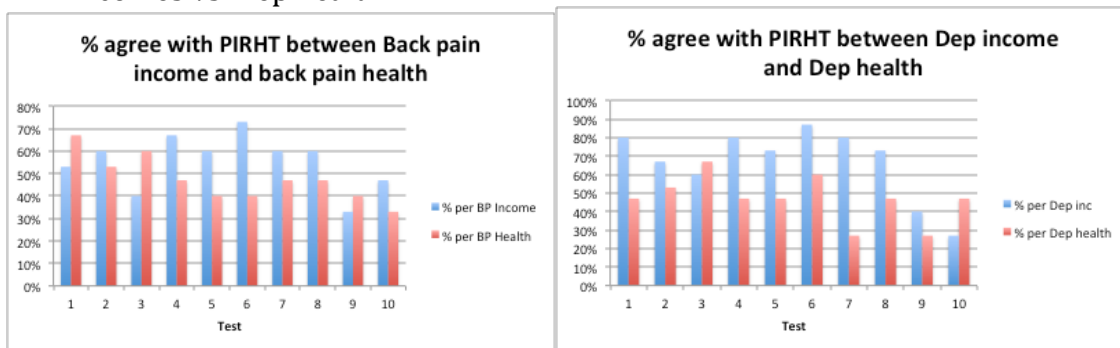

Table 9: T test for PIRHT for BP inc vs. BP health. Table 15: T test for PIRHT for DP inc vs. DP health

T test P- BP Inc vs BP healt T test DP inc vs DP health

Test #	P-value	Test #	P value
1	0.565536933	1	0.061675262
2	0.723952141	2	0.473374309
3	0.289303529	3	0.716456215
4	0.284919761	4	0.061675262
5	0.289303529	5	0.145829144
6	0.069284812	6	0.106761049
7	0.481439903	7	0.002363045
8	0.481439903	8	0.145829144
9	0.716456215	9	0.456070029
10	0.473374309	10	0.271320576

T test P-BP Inc Vs BP healt: P-value for two-sample independent t-test between back pain versions sorted by income and back pain version sorted by average duration of health.
T test DP inc vs DP health: P-value for two-sample independent t-test between depression version sorted by income and depression version sorted by average duration of health.

Is there a link between CIM & PIRHT questions that are very similar in terms of income group and/or average duration of health condition?

As both CIM and PIRHT deals with lowering inequality by reducing the degrees of correlation between the different dimensions of well-being, which in this case are income and average duration spent in particular health condition. Principle of income-related health transfer is similar to CIM, but in addition, it considers transfer of health from higher income persons to lower income persons as it decrease the correlation between income and average duration of health condition and the marginal distribution of health changes because of mean preserving health transfer. As a result, some tests of CIM and PIRHT are very similar in nature and it is interesting to see if there is some sort of link between the two concepts in terms of agreement rates.
In order to answer this question, a table was created through reexamining the questions of CIM and PIRHT and locating the questions of both principles that are very similar in terms of groups that are involved before & after the switch/health transfer as well as the average durations of condition pertaining to the switch/health transfer.

Overall (table 10 and figure 9), when comparing the similar tests of CIM and PIRHT, the agreement rates seem to be consistent with each other with on average from 0-

15% difference. It leads to the conclusion that when it comes to similar CIM and PIRHT questions, there are no notable differences in agreement rates with the exception of test 2, which has shown rather large differences (33% change). There does not seem to be a good explanation of the differences. After performing two-sample independent T test (table 11) between the similar CIM and PIRHT questions, it can be concluded that there seems to be no statistic significance between the agreement rates of both axioms. Even test 2 is found to be not significant. (p value=0.701450) Overall, subjects seem to be consistent of valuation of distributions of income and health for similar CIM and PIRHT tests.

Table 10: Agreement rates between similar CIM and PIRHT test

Whether there is a link between CIM and PIRHT questions that are very similar

			CIM				PIRHT		
	NR	% agree CIM	Groups involved	days in health	NR of Question	% agree with	Groups invo	days in healt	after trasfer
1	1	53%	1 Vs 2	20 Vs 17	20	38%	1 Vs 2	20 Vs 17	15 Vs 22
2	3	68%	2 Vs 3	13 Vs 17	18	35%	2 Vs 3	17 Vs 13	12 Vs 18
3	5	52%	2 Vs 4	8 Vs 17	16	57%	2 Vs 4	12 Vs 13	17 Vs 8
4	7	55%	2 Vs 4	20 Vs 8	14	53%	2 Vs 4	11 Vs 17	20 Vs 8
5	9	65%	4 Vs 5	5 Vs 8	12	65%	4 Vs 5	6 Vs 7	8 Vs 5
6	11	47%	3 Vs 5	17 Vs 5	10	55%	3 Vs 5	17 Vs 5	6 Vs 16
7	13	47%	1 Vs 4	20 Vs 8	8	60%	1 Vs 4	12 Vs 16	20 Vs 8
8	15	42%	1 Vs 5	20 Vs 5	6	57%	1 Vs 5	20 Vs 5	11 Vs 14
9	17	50%	1 Vs 3	13 Vs 20	4	58%	1 Vs 3	20 Vs 13	14 Vs 19
10	19	55%	3 Vs 4	5 Vs 13	2	62%	3 Vs 4	7 Vs 11	13 Vs 5

Figure 9: Proportion of subjects agreeing with each principle between similar CIM and PIRHT test
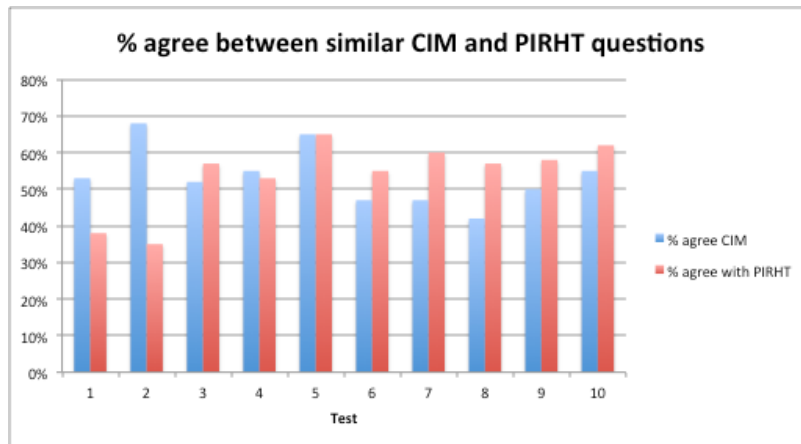
Table 11: T test between similar CIM and PIRHT question

T Test between similar CIM and PIRHT question

Test #	CIM question #	PIRHT question #	T test P value	T value
1	1	20	0.100797527	1.667086599
2	3	18	0.701450022	0.385233189
3	5	16	0.586287192	0.547226509
4	7	14	0.365456214	0.91204592
5	9	12	1	0
6	11	10	0.365456214	0.91204592
7	13	8	0.465396352	0.734759674
8	15	6	0.101951835	1.661333488
9	17	4	0.3638465	0.915133869
10	19	2	0.463105604	0.738554265

-

VIII. Discussion

The goal of this paper is to test empirically Correlation Increasing Majorization, one of Tsui's (1999) multidimensional inequality measures regarding the axioms behind correlation between different attributes and principle of income-related health transfer. It is important to find out if the general public agrees with Correlation Increasing Majorization and Principles of income-related health transfer behind correlation of different attributes. As the proportion of the subjects agreeing with CIM and PIRHT were calculated, overall, the majority of the respondents are inconclusive when it comes to agreement with decreasing correlation between different attributes. In CIM tests, the percentage of agreement is around 42% to 68% and most of the questions yield agreement rate fluctuates around 50%. After performing T-test, only two questions yield a significant difference from 50% and all other tests are not very significantly different from 50% at 5 % level. The reason behind testing the percentages of agreement rate against 50% is to see whether those percentages of agreement rate could due to indifference of the subjects or choice were purely random. The 50% difference is the minimum standard for each question to be considered satisfying the axioms being tested if the choice were not purely random. When it comes to PIRHT questions, majority of the questions yield agreement rate around 55% and only two tests scored lower than 40%. After computing the P-value from the T-test, only two questions have yielded significant difference from 50%; test with highest agreement rate and test with the lowest agreement rate.

Proportion of subjects agreeing with CIM and PIRHT are also compared with back pain and depression to see if the difference in agreement rate could due to different health condition. After performing two-sample independent T-test on back pain and depression of each question, it is confirmed that there is no major statistic difference in agreement rate between back pain version and depression version concerning both CIM and PIRHT except for one question of CIM (test 3) that shows slightly significance at p value=0.0507.

Sensitivity analysis is also done to check whether agreement with CIM and PIRHT depends on the ranking of the attributes. When it comes to CIM, the T-test shows there is no major significant difference in agreement rate between tested ranked by income versus test ranked by average numbers of days spent in back pain or with exception of only one test (test 7 on table 7), which has shown a significant difference at 10% level. Concerning depression versions, only two tests (test 1 and test 7 on table 8) has shown a significant difference at 10% level between the agreement rates for depression ranked by income and depression version of questionnaire ranked by average duration of health condition. The similar result is found with PIRHT tests with the back pain versions of the questionnaire; in which only one test (test 6 on table 9) is found to be significant at 10% level of difference in agreement rate between back pain sorted by income and back pain sorted by health and rest of the tests have not shown any significance. When it comes to health condition of depression, there are three tests (test 1, 4, and 7 from table 10),

which are to be considered statistically different at 10% level in terms of agreement rates between depression version of the questionnaire sorted by income and depression version of questionnaire sorted by average duration of health. There is no sensible conclusion that can be drawn from one significant test. When it comes to sensitivity analysis, it can be concluded that overall, when it comes to Correlation Increasing Majorization and Principle of income-related health transfer, majority of time agreement rates are affected by the sorting of the attributes such as income or average duration of health in this case.

Similar CIM and PIRHT tests are compared in terms of the agreement rate and to see if statistically there is any difference in the proportion of subjects agreeing with similar CIM and PIRHT. After performing a two-sample T-test, it has been confirmed that the possibility agreement rate of two similar CIM and PIRHT to be statistically significant different is slim.

Overall, it can be stated that people are consistent when it comes to valuing distribution of income and average number of days spent in either back pain or depression. Statistically speaking, different health condition does not make any difference in agreement rate across all respondents as well as the sorting of the different attributes.

IX. Conclusion

Although statistically, it is not very conclusive whether or not majority of the subjects agree with both Correlation Increasing Majorization and Principle of Income-related health transfer when it comes to measuring multidimensional inequality on distribution of two attributes: in this case income and quality of life, however, there are some limitation of this study that should be taken into consideration when evaluating the results. In this experiment, only two attributes were considered for measuring of inequality. As mentioned earlier, there are other attributes that are also crucial aspect for the measurement of well-being, for instance education attainment, and job opportunity. The findings of this study have not been tested for applicability for other attributes as mentioned above. Moreover, in this study, the inequality of health is measured through quality of life assuming constant life expectancy of 80 years old. What if the life expectancy varies as quality of life changes? Also, quality of life is assessed through a single morbidity (health condition) whether it's back pain or depression. In reality, people usually suffering from multiple health conditions (co-morbidities), which are not accounted for. All the above points could be incorporated in the future similar research on multidimensional inequality measures.

In this experiment, the subjects consist of university graduates and Master students that may raise concern that it is not adequate representative of general public. It is worth to mention that majority of the university graduates (23 out of 60) in this experiments are actually working in different health care field such as patient in-take, customer service of insurance companies. However, the size of that sample is less than half of the total sample size in the study. Nevertheless, there are published articles that observed no significant difference between the view of university students and those of general public. In a study of preference for health, de Wit, van Busschback, and de Charro (2000) concluded that there was no difference observed for preference of students compared to those of general public. In the article by Amiel and Cowell (1992), the authors argued that it might be better to use students than a sample for the general public when it comes to measurement of inequality. In their argument, the author believes that students are less biased when it comes to make judgments about inequality distributions and they are able to think analytically compared to the sample of general public, whom might have an intuition about what inequality is based on their possible experiences or specific situations which they may have come across rather than based on their impartial view based on what is considered fair defined by principle of distributive fairness.

A second concern maybe that the sample size of 60 questionnaires is not large enough to increase the explanatory power of the result. It maybe one of the reasons when T-tests are performed the P-value comes up not statistically significant from 50% agreement rates across both CIM and PIRHT questions. Two-sample independent T-tests are also performed on different versions of the questionnaire to see if the agreements rate with CIM and PIRHT depends on health condition or ordering of the attributes. Since each version of the questionnaire has only 15

respondents, therefore it may not have come up statistically significant for most of the tests.

Attempts were made in this paper to control for heuristics and biases that could affect subject's answer to each question. One of the attempts was to introduce two different health conditions such as back pain and depression to see how differ respondent supports the two axioms being tested. Each version of the questionnaire was either ranked by income or quality of life (average number of days spent in back pain or depression) again to see if sorting has made a difference in agreement rates across both principles. Within each version of the questionnaire, CIM questions and PIRHT questions were rearranged that made it hard for the respondent to spot any particular pattern. In addition, at each question, the policy that agrees with axiom being tested was also rearranged randomly to avoid heuristics. Although steps have taken to try to avoid heuristics and biases, some answers might be still prone to biases.

References

Abellan-Perpinan, J. M., Bleichrodt H., Pinto-Prades J. L., (2009) The Predictive Validty of Prospect Theory versus Expected Utility in Health Utility Measurement. Journal of Health Economics 28 (2009) 1039-1047.

Amiel, Y., Cowell F., 1992. Measurement of Income inequality: Experimental tests by questionnaire. Journal of Public Economics 47, 3-26.

Atkinson A. B., F. Bourguignon., 1982. The comparison of multi-dimensional distributions of economic status. Review of Economic Studies 49, 183-201.

Atkinson, A. B., 1970. On the measurement of inequality. Journal of Economic Theory 2, 244-263.

Atkinson, A. B., 1970. On the measurement of inequality. Journal of economic theory 2, 244-263

Bleichrodt, Han., Filko Martin., 2008. New Test of QALYs when health varies overtime. Journal of Health Economics 27 (2008) 1237-1249.

Bleichrodt, Han et al., 2005. Resolving Inconsistency in Utility Measurment under Risk: Test of Generalizations of Expected Utility. Management Science vol. 53. No. 3, March 2007, pp. 469-482.

Bleichrodt, Han., Pinto, J. L., Abellan-Perpinan, J. M., 2003. A Consistency test of the time trade-off. Journal of Health Economics 22 (2003) 1037-1052.

Bleichrodt, H., Rode K. I. IM., Van Ourti, T. 2009. Experimental tests of multidimensional inequality measures, mimeo, Rotterdam

Bleichrodt, H., Van Doorslaer E. K. A., 2006: A welfare economics foundation for health inequality measurement. Journal of Health Economics 25, 945-957

Boland P, Proschan F (1988) Multivariate arrangement increasing function with application in probability & statistics. J. Multivariate Analysis 25: 286-298.

Bourguignon, F. Chakravarty S. R., (2003) The measurement of Multidimensional Poverty. Journal of Economic Inequality 1, 25-49.

Carr-Hill, Ray A. (1989) Background Material for the workshop of QALYs. Social Science Med. Vol. 29, No.3, pp. 469-477, 1989.

De wit, G. A., Van Busschback J. T,. de Charro F. T. 2000. Sensitivity & Perspective in the valuation of health status. Health economics 9, 109-126.

Gdjdos, T., Weymark J., 2005. Multidimensional generalized Gini Indices. Economic Theory 26, 471-491.

Kolms, C., 1997. Multidimensional egalitarianism, Quarterly Journal of Economics. 91: 1-13.

Koshevoy, G., 1995. Multivariate Lorenz Majorizations. Social choice and Welfare 12: 93-102.

Koshevoy, G., 1998. The Lorenz Xonotope and Multivariate Majorization. Social Choice & Welfare 15: 1-14.

Koshevoy, G., K. Mosler., 1996. The Lorenz Zonoid of a multivariate distributions. Journal of the American Statistical Association 91 (434): 873-882

Le Breton M., 1986. Essaos Sir ;es fpmde,emts de l'analyse economique d' l'inegalite. These pour le Doctorat d' etat. Universite de Rennes I.

Moyes P., 1999. Comparisons of heterogeneous distribution & dominance criteria. Economies et prevision. 138-139, 225-146 (In French)

Nord, Erik., 1990. EuroQol: Health-related Quality of Life Measurement. Valuation of health states by the general public in Norway. Health Policy, 18 (1991) 25-36.

Organisation For Economic CO-operation and development (OECD 2008). Annual Report 2008. Paris, France.

Tsul, K. –y. 1995. Multidimensional Generalizations of the Relative & absolute inequality indices: The Atkinson-Kolm-Sen approach. Journal of Economic Theory 67, 251-265.

Tsui, K. –y1999. Multidimensional inequality & multidimensional generalized entropy measures: An axiomatic derivation. Social Choice & Welfare 16, 145-157.

Van Asselt, ADI. et al., 2009. The EQ-5D: A useful quality of life measure in borderline personality disorder. European Psychiatry 24 (2009) 79-85.

Appendix A (Back pain income version)

Thank you very much for participating in this survey. The purpose of this survey is to obtain information on how people value distributions of income and average number of days in a year spent in severe back pain. There are no 'right' or 'wrong' answers in this questionnaire. We are only interested in your personal opinion.

Imagine a country Alfaland, in which two governmental policies have been proposed. Both policies have identical effects on the country, except for the income and average of days in a year spent in severe back pain.

The population of Alfaland consists of 5 groups of people. All 5 groups 1) are equally large in size, 2) have equal life expectancy of 80 years and 3) have the same average age. All individuals experience severe back pain (at least once) in their lives and the difference between the groups is the average duration of severe back pain which is described as average number of days per year. On all days without severe back pain, the individuals have perfect health. Within every group all persons have the same income and same number of days with severe back pain.

In the following table, the five numbered columns represent the five groups of the people in that country.

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	13	8	5

The five columns, named 1, 2, 3, 4, and 5, give the income and average number of days in a year spent in severe back pain for each group. Income is the net monthly income in Euros per person. The price level in Alfaland will be similar to the price level in the Netherlands in 2010. The average number of days in severe back pain do not affect the corresponding income. Underneath, we provide a description of severe back pain.

Severe back pain
People with severe back pain have no problem walking and moving around on their own; there is no assistance necessary for mobility. They neither need assistance to wash, dress or take care of themselves. However, severe back pain affects performance in usual activities: e.g. they experience some problems in sitting down for a long period of time or in studying without taking some breaks in between. Also they have some problem doing housework such as cleaning and doing dishes for a long period of time. In addition, people experiencing severe back pain have a

moderate degree of pain or discomfort: e.g. the pain can be a dull ache, a sharp or piercing or burning feeling and sometimes it could have other symptoms such as weakness, numbness or tingling feeling. There is no psychological effects associated with individuals with severe back pain; they are not anxious or depressed.

In each question below, there are two policies which will affect Alfaland in different ways in terms of distributions of income and average number of days of severe back pain spent in a given year. You are asked to choose repeatedly between the two distributions of income and average number of days of severe back pain per year which are labeled under Policy A and Policy B. Please indicate the policies that you pick by highlighting them.

1.

Policy A

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	13	8	5

Policy B

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	17	20	13	8	5

2.

Policy A

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	7	11	5

Policy B

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	13	5	5

3.

Policy A

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	13	17	8	5

Policy B

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	13	8	5

4.
Policy A

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	13	5	5

Policy B

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	14	17	19	5	5

5.

Policy A

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	8	13	17	5

Policy B

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	13	8	5

6.

Policy A

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	13	5	5

Policy B

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	11	17	13	5	14

7.
Policy A

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	20	13	8	5

Policy B

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	8	13	20	5

8.
Policy A

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	12	17	8	16	5

Policy B

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	8	8	5

9.
Policy A

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	20	13	5	8

Policy B

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	20	13	8	5

10.
Policy A

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	17	17	5

Policy B

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	6	17	16

11.
Policy A

	1	**2**	**3**	**4**	**5**
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	17	17	5

Policy B

	1	**2**	**3**	**4**	**5**
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	5	17	17

12.
Policy A

	1	**2**	**3**	**4**	**5**
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	20	13	6	7

Policy B

	1	**2**	**3**	**4**	**5**
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	20	13	8	5

13.

Policy A

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	8	8	5

Policy B

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	8	17	8	20	5

14.
Policy A

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	11	13	17	5

Policy B

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	20	13	8	5

15.
Policy A

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	13	5	5

Policy B

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	5	17	13	5	20

16.

Policy A

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	12	13	13	5

Policy B

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	13	8	5

17.

Policy A

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	13	17	20	5	5

Policy B

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	13	5	5

18.

Policy A

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	13	8	5

Policy B

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	12	18	8	5

19.

Policy A

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	5	13	5

Policy B

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	13	5	5

20.

Policy A

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	20	17	13	8	5

Policy B

	1	2	3	4	5
Income (Euro)	550	780	980	1250	2350
Average number of days of severe back pain per year	15	22	13	8	5