uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Jianwei Zhong

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Economics)

GRADE / DEGREE

Department of Economics

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Three Essays on Population Health

TITRE DE LA THÈSE / TITLE OF THESIS

Victoria Barham

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

Rose Anne Devlin

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

Paul Makdissi

Marcel Mérette

Quentin Wodon (World Bank – Washington, DC)

Stanley Winer

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Three Essays on Population Health

A thesis presented

by

Jianwei Zhong

to

Faculty of Graduate and Postdoctoral Studies

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

University of Ottawa

Ottawa, Ontario

January 2010

Your file Votre référence
ISBN: 978-0-494-66250-2
Our file Notre référence
ISBN: 978-0-494-66250-2

To my father Zhenguang Zhong, my mother Xiuzhen Gan, and my wife Yu Lan

Abstract

This thesis consists of three chapters that focus on the problem of measuring population health.

Chapter One: Distribution-Sensitive Measurement of Population Health. It starts from the observation that standard measures of population health aggregate individual's well-being without taking account of any inequality in the distribution of health in the population. It tries to fill this gap in the literature by exploring two approaches to constructing a distribution-sensitive measure of population health. The first approach, based on Sen's (1976) poverty index, provides an operationally-simple method for comparing the health of two populations when these populations vary with respect to a number of different health attributes. The second approach, drawing on work by Davidson and Duclos (2000), demonstrates that stochastic dominance can be used to compare the health of populations, both when a cardinal measure of individual well-being is available, and when individual health states are described only by a multi-dimensional vector of health attributes.

Chapter Two: Robust distributionally-sensitive comparisons of population health. Drawing on the literature on the measurement of poverty and income inequality, it considers the use of stochastic dominance in measuring population health and compares this with a new measure developed in this thesis, the Mean Inequality Measure. It shows that stochastic dominance may not be the most appropriate approach to the measurement of population health, largely because of its

sensitivity to individuals in very poor health. These two approaches are contrasted empirically, by applying them to the evaluation of health data from the US and Canada.

Chapter Three: Measurement of Population Health Over Time. It extends the study of distribution sensitive measures of population health to take account of the specific conceptual challenges posed by the fact that individual health outcomes change over time. The strategy to tackle this issue in this paper is to first establish a number of axioms which should be satisfied by any useful measure of population health over time, and second to develop a measure which satisfies these properties. Subsequently, this measure is compared to a number of other possible approaches.

Acknowledgments

I am deeply grateful to my supervisors, Professor Rose Anne Devlin and Professor Vicky Barham. They greatly enlightened me in the path of my exploration for the economic truth. And they selfishlessly tutored me to think creatively and to express rigorously in both theoretical and empirical levels. I will not forget all the experiences and tips I obtained when I worked with them. And I will not forget their excellent guidances and enormous encouragements during the long journey of finishing this thesis. Besides, I was lucky enough to benefit a lot from the life wisdom they shared with me. In my eyes, they are always optimistic, diligent, patient and full of curiosity about the world. I am proud of being their student.

I would like to thank all the members of my thesis committee, Professor Stanley Winer, Professor Marcel Mérette, Professor Catherine Deri Armstrong, Professor Paul Makdissi, Professor Quentin Wodon, Professor Gamal Atallah for their highly valuable and constructive comments. In particular, I benefited a lot from the course taught by Professor Paul Makdissi, from which I got systematic knowledge on the theories of povety measurement and income inequality measurement.

I would like to thank Dr. William Hogg , and Simon Dahrouge, Dr. Meltem Tuna, Dr. Grant Russell, Dr. Kevin Pottie, Dr. Clare Liddy, Dr. Sharon Johnston Dr. Marie-Hélène Chomienne and all other colleagues in C.T.Lamomnt Centre of

Élisabeth Bruyère Research Institute for their great supports and encouragments during the final stage of my thesis writing.

I would like to thank Diane Ritchot, Irène Paré, Bernadette Benoit, Luce Laviolette and all other administrative personnel at the Department of Economics, for their helps and supports during my study periods at the department.

My full gratitude go to my parents and parents-in-law, elder brother and sister-in-law, and my dear wife. Their love gave me endless power to solve the difficulties one after another in my ph.d study and in my daily life. I am also indebted to my friends who have helped me to go through the past five years, especially XiaoJing Wu, Guohan Zhu, Li Xue, Xiaochuan Wang, Rui Hao, Ting Xu, Minhao Ren, Olga Milliken, and Reza Ghazal.

Ottawa, Canada, May 2009

Contents

List of Tables

List of Figures

Legend

ALD	Absolute Lorenz Dominance
AMIM	Average MIM
CM	Conventional measure
DALY	Disability adjusted life year
DMIM	Dynamic version of MIM
HALY	Health adjusted life year
HUI	Health utility index
LD	Lorenz Dominance
MCM	Modified conventional measure
MIM	Mean Inequality Measure
LD	Lorenz Dominance
SD	Stochastic Dominance
SRH	Self reported health
TMIM	Time version of MIM

Introduction

There is an increasing understanding that a "good " measure of population health is a crucial tool if health policy makers are to be able to efficiently allocate limited health resources in order to improve the population health. Over the past two decades, a number of different paths have been followed in an attempt to establish useful approaches to meaningfully measuring population health. Despite these efforts, there remains considerable ambiguity with respect to how to evaluate and determine whether a measure itself is a "good" measure of population health.

Existing methodologies more or less fall into three main streams. The first stream focuses on measuring the central tendency of the health of the population. Examples of such summary measures are life expectancy, HALYs, QALYs, DALYs, etc. These measures rank the health of different populations primarily by comparing differences in the mean or median health status of individuals. In constrast, they pay little heed to the second moment of the distribution of health experience in the population. Therefore, in those cases where the distribution of population health has a skewed or scattered pattern (which the empirical evidence suggests may be very common) these measures will generally fail to take account of the differences between the health of different populations with comparable means but very different variances, and may result in inappropriate policy choices.

In contrast, the second stream gives much more weight to the distribution of the health of individuals in the population. This stream includes measures such as the Gini coefficient, and the mean logarithmic variation, and calculates a measure of dispersion based on the

relative location of individuals' health levels in the health distribution. However, these measures typically are insensitive to changes in the absolute health levels of the population, as opposed to changes in the distribution of health. This is problematic when considering many possible public policy interventions: for instance, they are liable to rank as equally attractive a proposal to bolster the usage of health services by people in poor health, and a competing proposal which would cut the health resources available to already healthy invididuals. Whereas both proposals would reduce inequality in the distribution of health, it is clear that increase the use of health services by the poor will lead to a higher average level of individual health, and should therefore be the preferred policy strategy.

Finally, the third stream views the health of the population as an aggregation of measures of individuals health, but uses unequal weights for different individuals. Individual weights are typically determined according to various functions that are dependent on the health levels of individuals. Atkinson's index, Theil's index, and rank-dependent QALYs are typical examples. And a variant of this stream is the sequential dominance approach, which is based on sequential comparisons of the health of subpopulations, generated from Lorenz curves, concentration curves or stochastic dominance curves. These approaches incorporate a concern for equity into the measurement of population health by giving higher weights to less healthy individuals, which brings a new perspective to the policy making exercise. Unfortunately, however, the weights assigned to each individual are implict and it is hard to determine whether the weights incorporated into these measures are consistent with the preferences of the health policy maker.

Different welfare functions are employed to justify the use of each of the above methodologies. However, there is a general lack of clarity regarding the strengths and weaknesses of each approach and, more fundamentally, of the desired features of a measure of population health that would be a meaningful tool for policymakers in real-world settings. To this end, it is important to clarify the key concerns and preferences of health policy makers when faced with making decisions over different policy alternatives. Essentially, the objective of health policy makers can be summarized as seeking to improve the overall health level of the population while also reducing inequality in the distribution of health across individuals. Given this observation, this thesis seeks to contribute to our understanding of what constitutes a "good" measure of population health, and also proposes two new measures. In Chapter One, the difficulties of comparing the health of different populations are examined under different hypothetical senarios. These scenarios lead to the identification of five criteria that would need to be met by any satisfactory measure of population health. The criteria of monotonicity and distribution sensitivity included in the five proposed criteria embody the two core goals of health policy makers, and the mean-inequality tradeoff criterion helps to define the relationship between these two goals. The other two criteria (completeness and convexity) ensure that the measure of population health is always able to rank different alternatives, and that it can be used in multiattribute settings. It is demonstrated, however, that most of the existing approaches to measuring population health fail to satisfy all of these crtieria - in particular, because they fail to meet at least the mean-inequality tradeoff criterion. (The one major exception is the stochastic dominance approach.) In view of the weaknesses with these existing approaches, I pro-

pose a new measure – the Mean-Inequality Measure (MIM) of population health - which (along with the stochastic dominance approach) satisfies all five criteria, and can be be operationalized in different settings.

Subsequently, in Chapter Two, the focus is on comparing stochastic dominance and the Mean-Inequality Measure as approaches for the measurement of population health. The primary concern is to illustrate the differences in the capacity of these two measures to incorporate a concern for equity into the measurement of population health. A strength of the MIM measure is that it has the ability to explicitly capture the preferences of the policy maker with respect to trading off average health and health inequality in measuring the health of the population. In contrast, it is shown that the stochastic dominance approach assigns very high weight to the least-well-off individual in the population, and consequently in empirical applications it can generate very counter intuitive rankings: for example, when comparing a health distribution where all individuals have high health levels except for one individual who has a very low health status, versus another health distribution where all individuals have medium health levels, it can be shown that the first distribution is always dominated by the latter at a certain order of stochastic dominance and hence gain priority in receiving health resources. This issue is of particular concern when measuring population health because there is good reason to expect that the health of the least-healthy individuals in the population (e.g., those who are born with serious congenital health problems) may not be significantly improved even if essentially unlimited health care dollars are directed to caring for these individuals. Therefore, it is argued that when policy makers are concerned with generating rankings over the health distributions of different populations, and

particularly where some flexibility with respect to equity considerations is desired, MIM is a better candidate.

Finally, in Chapter Three, some initial steps are taken to establish how to extend the MIM measure to a dynamic setting, a crucial objective if health policy makers are to develop optimal long term population health strategies. Drawing on the literature of poverty measurement and income inequality measurement, a number of axioms are identified which must be exemplified by a satisfactory dynamic meaure of population health. The five criteria articulated in Chapter One are in effect appropriately reformulated as five axioms (monotonicity axiom, replication invariance axiom, weak anonymity axiom, transfer axiom and decomposability axiom) which reflect specific concerns with respect to the measurement of population health over time, and proposes a potential candidate measure - the Dynamic Mean-Inequality Measure (DMIM) - which satisfies these five axioms. Although this chapter does not propose a full axiomatisation of the DMIM, it shown that this measure is well-designed to capture the sequence of health experience of each individual in the population over time, and that this distinguishes it from a number of other possible approaches to measuring population health in a dynamic setting. The differences between the $DMIM$ and some of the potential alternative measures are also illustrated by the application of a longitudinal health survey data from China.

There are two possible extensions following the study in this thesis: one is to fully axiomatize the DMIM, the other is to measure the specific policy effects to population health through the application of DMIM. These efforts will deepen the understanding towards the strengths and weaknesses of different techniques to measuring population health in an over

time setting, and may provide useful guidance for the health policy maker to reach more effective decision in allocating health resources from a long term perspective.

Chapter 1
Distribution-Sensitive Measurement of Population Health

1.1 Introduction

The importance of designing health policies to promote the health of populations rather than of individuals is widely recogized in health policy circles. Public sector committment to developing policy based on a population health framework is reflected in such documents as The Population Health Template (Health Canada, 2001), the Verona Benchmark (WHO/Euro); and the Population Health Investment Policies of the Working Party on Social Policy of the OECD. These international efforts to reframe health policy to focus on strategies that most effectively promote improvement in the health of populations reflects a general understanding that individual health outcomes are broadly determined, including by socio-economic factors, and not merely by the quality of health care services and individual variation in innate health capital. Moreover, the population health approach emphasizes that good health care policy is health care policy which attenuates inequality in health outcomes.

Strikingly, however, existing methodologies for measuring the health of populations are not yet truly able to meaningfully capture information about disparaties in the distribution of health across individuals. Various indicators of the health of populations are proposed, but they generally aggregate information about the health of individuals without

taking account of inequality in the distribution of health. Also, it is worth to note that there is a growing (and already signficant) literature on health inequality, and on the relationship between socio-economic inequality and health inequality. However, the focus of this literature has been on the problem of inequality, including the sources of this inequality, and there has been scant attention paid to the issue of integrating concerns about the distribution of health in the measurement of the health of the population.The importance of this problem for policymakers is evident.

The benefit to policymakers of developing a means to meaningfully aggregate information about individual health, and about health inequality, in a single summary measure of population health is clear. For example, in determining how to allocate limited health care resources, it is crucial that policymakers be able to determine whether it is preferable to choose a policy which has a widespread albeit shallow impact (for example, national flu shots) versus investing in health care services which will have a very significant impact on the lives of a far smaller sub-sector of the population (for example, financing a number of hip replacements). Or, it may be important to decide whether to allocate more resources to a disease which appears to strike members of the community relatively randomly, or whether to put additional effort into fighting diseases which most often afflict individuals who are also struggling with other health problems, e.g., diabetes and Aboriginal Canadians. In the international context, it is important for organisations such as World Health Organization to know whether it is more useful to undertake a health investment in a country with a relatively low average level of health, but where disparity in health outcomes

is small, or in a country that may have a higher level of average health, but much greater variation in individual health outcomes.

This paper examines the problem of developing a distributionally-sensitive measure of population health. To identify some of the challenges that must be met in developing an operationally-meaningful population health measure that is sensitive to health inequality, we first consider a number of different hypothetical populations, and use the analysis of these hypothetical cases to distill five key criteria that must be met by a distributionally-sensitive measure of population health. Next, several widely-used measures of population health are evaluated according to these five criteria. For illustrative purpose, they are categorized into three main groups, and a typical indicator from each group (HALY, DALY and HUI3) is scrutinized according to the five criteria. We show that none of the existing indicators meet these criteria and demonstrate that the principle weakness of these measures is their lack of sensitivity to information about the distribution of health across individuals. Two different approaches to constructing a distributionally-sensitive measure of population health are then proposed. The first approach, which is relatively straightforward to operationalize, is inspired by Sen's (Sen, 1976) poverty index, and combines data about the mean and inequality in the health of individuals in a way that enables comparisons to be drawn between the health of two populations when these populations vary with respect to a number of different health attributes. It can also be extented to a complex form that incorporates the substitution effect between the mean and inequality of the population health. A limitation of this approach is that it would be sensitive to the choice of inequality index in extreme case. The second approach is based on the knowledge of stochastic dominance,

and it is able to be implemented both when a cardinal measure of individual well-being is available, and when individual health states are described only by a multi-dimensional vector of health attributes. The last section concludes.

1.2 Comparing the Health of Populations: Some Examples

This section provides several examples in order to illustrate a number of practical challenges that must be met by a measure of population health that is fully sensitive to distributional considerations. This exercise will enable us to identify a number of key properties that must be embodied in a health inequality-sensitive measure of population health. For the sake of discussion, in each of the examples below we consider small populations of only 7 individuals, and two populations, A and B.

Example One: Mean and Variance

Our first, and simplest example, illustrates the importance of incorporating some measure of health inequality in the measurement of population health. In the table below, h denotes the health level of a specific individual, and we assume that $h_i > h_j$ if $i < j$.

Table 1 individual	Population A	Population B
1	h_3	h_5
2	h_4	h_5
3	h_5	h_5
4	h_5	h_5
5	h_5	h_5
6	h_6	h_5
7	h_7	h_5

Casual inspection reveals that some individuals in population A are healthier than others, whereas all individuals in population B have the same level of health. Moreover,

the mean health level is the same: h_5. Intuitively, one has the sense that these populations should not be judged to be equally healthy: the fact that health is so unequally distributed in population A should matter. However, if the measurement of the health of the population is simply the arithmatic mean of the health of the individuals in that population, then both of these populations will be considered to be equally healthy. Nevertheless, if one takes the view that the distribution of health matters, then in comparing these two populations it is not obvious which one should be judged healthier. On the one hand, population A has some truly healthy individuals (individuals 1 and 2) whereas in population B everyone is in "middling health". On the other hand, health is more equally distributed in population B than in population A and so a measure of population health that traded off increases in the mean against increases in the variance would tend to rank population B as having higher health.

Another issue that this example illustrates is the problem of 'transferring health' between individuals[1]. In population A, it might be possible to transfer resources from healthier individuals to less healthy ones. Consider a transfer which shifts one unit of health from individual 1 (the healthiest individual) to individudal 2 (the second most healthy individual). Should this transfer increase the measured level of population health by as much as a transfer of one unit of health from individual 1 to individual 7, the least healthy individual? Or from individual 6 to individual 7? In the literature on the economics of in-

[1] In practice, as noted by Dachs (2002), it is not obvious that health can be transferred between individuals in the same way that income can, for instance. However, this does mean that the Pigou-Dalton principle is not without interest in developing meaningful summary measures of population health. When considering the implementation of two different health policies, the ultimate impact of those policies may be to ultimately distribute health differently across members of the population. The Pigou-Dalton principle in effect gives us a tool for comparing the level of population health in the two different final distributions.

come inequality, an income inequality measure is said to satisfy the Pigou-Dalton Principle if equality-enhancing transfers are weighted more heavily when resources are transferred from individuals who are richer than average to individuals who are poorer than average; additionally, a measure is said to have transfer sensitivity if transfers which reduce inequality amongst the poor (but which are from one poor person to another) are weighted more heavily than transfers which reduce inequality amongst the rich. It seems reasonable to require that a distribution-sensitive measure of population health should satisfy the Pigou-Dalton principle and the principle of transfer sensitivity.

Example 2: Multiattribute Individual Health

Our second example deals with population health measurement when individual health is multidimensional. We can think of indviduals having both physical (h_1) and emotional health (h_2). Once again, we assume that $h_{i,j} > h_{i,k}$ if $j < k$.

Table 2A	Population A		Population B	
individual	h_1	h_2	h_1	h_2
1	$h_{1,2}$	$h_{2,8}$	$h_{1,3}$	$h_{2,5}$
2	$h_{1,2}$	$h_{2,8}$	$h_{1,4}$	$h_{2,6}$
3	$h_{1,2}$	$h_{2,8}$	$h_{1,3}$	$h_{2,7}$
4	$h_{1,2}$	$h_{2,8}$	$h_{1,5}$	$h_{2,8}$
5	$h_{1,2}$	$h_{2,8}$	$h_{1,2}$	$h_{2,4}$
6	$h_{1,2}$	$h_{2,8}$	$h_{1,5}$	$h_{2,3}$
7	$h_{1,2}$	$h_{2,8}$	$h_{1,6}$	$h_{2,7}$

Table 2B	Population A		Population B	
individual	h_1	h_2	h_1	h_2
1	$h_{1,2}$	$h_{2,8}$	$h_{1,3}$	$h_{2,7}$
2	$h_{1,2}$	$h_{2,8}$	$h_{1,4}$	$h_{2,8}$
3	$h_{1,2}$	$h_{2,8}$	$h_{1,3}$	$h_{2,5}$
4	$h_{1,2}$	$h_{2,8}$	$h_{1,5}$	$h_{2,6}$
5	$h_{1,2}$	$h_{2,8}$	$h_{1,2}$	$h_{2,7}$
6	$h_{1,2}$	$h_{2,8}$	$h_{1,5}$	$h_{2,3}$
7	$h_{1,2}$	$h_{2,8}$	$h_{1,6}$	$h_{2,4}$

Looking at Table 2A, and examining population A, we notice that all individuals have a very high health level of attribute h_1, and a very low level of attribute h_2. In contrast, individuals in population B differ in health levels in the two attributes, for attribute h_1,they all have lower health levels than in population A, while for attribute h_2 they all have higher

health levels than does any individual in population A. A natural first question brought out by this example is whether or not one of the attributes is more important? Or should they be equally weighted? The answer to this question makes a big difference. For example, if $h_1$is far more important, then population A should be judged to be healthier, but if h_2 is very important, then the ranking would be reversed. This demonstrates that a summary measure of population health which puts equal weight on all attributes may be very misleading: there has to be careful thought put into the appropriate weighting of the different dimensions of an individual's health status.

A second issue that this example raises is the possibility that the burden of any one health problem may be either positively or negatively affected by the existence of other health problems. For example, individual 5 has a health level of $h_{1,2}$ in both populations A and B. But in population A, this individual has attribute $h_{2,8}$ whereas in population B this individual has attribute $h_{2,4}$. If individual health status is not a strictly additive function of their health status for each attribute, but rather on the interaction between the attributes, then the difference between the level of the health of these two individuals should not simply be equal to the difference between $h_{2,8}$ and $h_{2,4}$. This is important when we compare Tables 2A and 2B: notice that the difference between these two tables is simply in the distribution of the second attribute in population B. If the interaction between attributes affects individual health, then population B should not be equally healthy in both of these possible states of the world, although the distribution of each attribute, viewed individually, across the population is the same. In summary, a satisfatory measure of population health must reflect appropriately the multidimensional nature of individual health states.

Example Three: Changes in the Distribution

Our third example once again considers two populations with multidimensional health, and illustrates some of the challenges encountered in comparing the impact of different health policies on the health of a given population. Comparing population A and population B with respect to the distribution of health in the original state, we note that Policy A - for example, a compulsory flu vaccination - results in an outcome whereby all individuals in population A achieve a health outcome equal to the highest possible level for attribute h_1 while maintaining the original levels for attribute h_2. In contrast, the implementation of Policy B - complex drug therapy for individuals suffering from cardiovascular disease - leads to a situation where in population B individuals 1 to 4 simultaneously increase their health to a higher level for attribute h_2 while keeping the same health levels for attribute h_1 and there is no change in either attribute for individuals 5 through 7.

Table 3	Population A		Population B		Original health	
individual	h_1	h_2	h_1	h_2	h_1	h_2
1	$h_{1,1}$	$h_{2,9}$	$h_{1,3}$	$h_{2,6}$	$h_{1,3}$	$h_{2,9}$
2	$h_{1,1}$	$h_{2,9}$	$h_{1,3}$	$h_{2,6}$	$h_{1,3}$	$h_{2,9}$
3	$h_{1,1}$	$h_{2,8}$	$h_{1,3}$	$h_{2,6}$	$h_{1,3}$	$h_{2,8}$
4	$h_{1,1}$	$h_{2,7}$	$h_{1,2}$	$h_{2,6}$	$h_{1,2}$	$h_{2,7}$
5	$h_{1,1}$	$h_{2,6}$	$h_{1,2}$	$h_{2,6}$	$h_{1,2}$	$h_{2,6}$
6	$h_{1,1}$	$h_{2,5}$	$h_{1,2}$	$h_{2,5}$	$h_{1,2}$	$h_{2,5}$
7	$h_{1,1}$	$h_{2,4}$	$h_{1,2}$	$h_{2,4}$	$h_{1,2}$	$h_{2,4}$

Clearly, both policies have improved population health. But which delivers more "bang for the buck"? Even if each attribute were equally important, is it better to have a relatively modest increase in the health of all individuals, or a more significant improvement in the health of a select few? Is it better to "cure" certain conditions, or to simply "ameliorate"? Is it better to reduce inequality along dimensions of health where there are

very significant differences in individual health, or should we attach equal importance to attacking health problems which are "mild" but relatively evenly distributed in the population?

1.3 Criteria for Evaluating Competing Measures of Population Health

In view of the above examples, it is useful to identify some key properties which must be satisfied if the population health measure is to appropriately reflect the related major concerns. It is useful to introduce some notation in order to precisely describe the properties that should be exhibited by a more satisfactory measure of population health.

Suppose we have a population of n individuals. The health of the population can then be described as a vector of health states for each of the n individuals, i.e. $P = (x_1, x_2, ..., x_i, ..., x_n)$, where $x_i = (x_{i1}, x_{i2}, ..., x_{ij})$ denotes the health state vector for individual i, and $x_i \in \Omega$, where $\Omega = X_1 \times X_2 \times ... \times X_j$, is the j dimensional health state space. Observe that Ω is closed and convex, and $X_1 ... X_j$ are subsets of Ω respectively, where each X_i corresponds to one health attribute. The health of the population is then a measure $M(P)$, and this measure should exhibit the following desirable features.

1. $M(P)$ provides a complete ranking: it ranks the health states with respect to each attribute and in the multi-dimensional space $M(P)$ identifies both best and worst health states. This requires that $M(P)$ is able to capture and reflect a specific "reasonable" order relation over the health states. Essencially in the view of $M(P)$, each X_j is a totally ordered set, and consequently the upper bound and lower bound

are located as $\overline{x_j}$ and $\underline{x_j}$. For each attribute, it is possible to apply the order relation "$\succeq$" over $(x_{1j}, x_{2j}, ..., x_{ij}, ..., x_{nj})$. And "$\succeq$" also could be defined for Ω, with maximal element $(\overline{x_1}, \overline{x_2}, ..., \overline{x_j})$ and minimal element $(\underline{x_1}, \underline{x_2}, ..., \underline{x_j})$ respectively being the best health state and the worst health state. The order relation is the basis for obtaining numeric depictions for the multi-dimensional health states.

2. Monotonicity. The measure of population health must be Pareto-sensitive. Consider two populations $P = (x_1, x_2, ..., x_i, ..., x_n), P' = (x'_1, x'_2, ..., x'_i, ..., x'_n)$, where $x_i \succeq x_j$ (respectively, $x'_i \succeq x'_j$) if $i > j$. Then if $x_i \succeq x'_i$ for all i,and $x_i \succ x'_i$ for some i, it must be true that $M(P) > M(P')$.If in comparing two populations every individual in the reference population is at least as healthy as their comparator in the alternative population and at least one individual is healthier than his comparator, then the first population is healthier.

3. The combination of health attributes for each individual matters. Consider four populations $P = (x_1, x_2, ..., x_i, ..., x_n), P' = (x'_1, x'_2, ..., x'_i, ..., x'_n), P'' = (x''_1, x''_2, ..., x''_i, ..., x''_n), P''' = (x'''_1, x'''_2, ..., x'''_i, ..., x''''_n)$ where $x_i \equiv x'_i \equiv x''_i \equiv x'''_i$for all $i \neq k$. $x_k = (x_{k1}, x_{k2}, ..., x_{kj})$, $x'_k = (x'_{l1}, x'_{l2}, ..., x'_{lj})$, $x''_k = (x''_{l1}, x''_{l2}, ..., x''_{lj})$ and $x'''_k = (x'''_{l1}, x'''_{l2}, ..., x'''_{lj})$. Suppose $x_{kj} = x'_{kj} = x''_{kj} = x'''_{kj}$ for all $j \neq l, m$. Without loss of generality, suppose that $x_{kl} = 0, x_{km} = 0, x'_{kl} = 0, x'_{km} = \widehat{x}, x''_{kl} = \widehat{x}, x''_{km} = 0$ and $x'''_{kl} = \widehat{x}, x'''_{km} = \widehat{x}$. Then $(M(P') - M(P)) + (M(P'') - M(P)) \leq M(P''') - M(P)$. In other words, the sum of magnitude in the increase in the health of the population when a given individual's health improves along one dimension (l) plus when it

increases along some other dimension (m) is less than the magnitude of the increase in the health of the population when there is a simultaneous (but equivalent) increase along both dimensions.

4. The distribution of health across individuals matters. It is not equivalent to have a population in which all individuals have the same average level of health, and a population with the same mean level of health, but in which half the population is very healthy (healthier than the mean) and half of the population is very unhealth (less healthy than the mean). Formally, let F be the cumulative health distribution function across individuals' health, then the measure system is modified to be $M(P, F(P))$, $M(P', F(P'))$.If average health in P = average health in P' then $M(P, F(P))$ $=M'(P', F'(P'))\ iff\ P = P'$.

5. A reduction in inequality will compensate for a small loss in average health. Consider 2 populations $P = (x_1, x_2, ..., x_i, ..., x_n), P' = (x'_1, x'_2, ..., x'_i, ..., x'_n)$ where $x_k \leq x_{k+1}$ and $x_k = x'_k$ for $2 \leq k < n$. Let $x_1 = x'_1 + \epsilon_1, x_n = x'_n - \epsilon_2$ where $\epsilon_1 < \epsilon_2$. Then for small enough ϵ_1, ϵ_2, $M(P) > M(P')$. Observe that a corollary of this property is the Pigou-Dalton principle: Consider two populations with the same average level of health and which are identical except for individuals A and B. In both populations, individual A is healthier than individual B; however, in population 2 the health of individual A is x units less than in population 1 and the health of individual B is x units greater in population 2 than in popluation 1. Then population 2 is healthier than population 1. If $P = (x_1, x_2, ..., x_i, ..., x_n)$ with the ranking over individuals as

$x_1 \succeq x_2 \succeq ... \succeq x_n$, and $P' = (x_1, x_2 - \Delta_2, ..., x_i + \Delta_i, ..., x_n)$, where $\sum_{k=1}^{n} x_k = \sum_{k=1}^{n} x'_k$ then $M(P) < M(P')$. It is worth underscoring the fact that the Pigou-Dalton does not involve literal transfers of health from one individual to another, but merely comparisons of populations which are similar except for a reduced disparity in the distribution of health across each population.

These five criteria, although directly distilled from the hypothetical examples we described above to fit in the health measurement context, could also be identified in exact or similar terms in the literature to characterize preferences and utility functions. For example, the completness is one of the necessary conditions for a preference to be rational. Non-satiation is an important assumption for a well behaved utility function, which is syncronym to monotoniciy in our health meaurement setting. And the convexity mentioned here is also a concept to define the conditional utility functions in the multiattribute setting. The last two criteria reflect the concerns on the relationship between average pattern and precise pattern and their tradeoff, which are missing or insufficiently captured in the ordinary social preference. Consequently, it is reasonable to combine these five criteria to evaluate the appropriateness of a measure of population health.

1.4 Evaluating Existing Approaches for Comparing Population Health

In view of our discussion above, it is useful to review existing approaches to the measurement of population health to establish whether or not they satisfy the criteria laid out above.

Given that the literature on the measurement of population health is very extensive, it is not possible to undertake an exhaustive comparison. Consequently, our review in Table 4 first groups measures into three broad categories, and then examines specific examples in each of these categories.

For illustrative purposes, it is useful to examine one example of each of these three broad approaches in greater detail. A particularly influential health expectancy measure has been the HALY(health adjusted life years)(Molla *et al*, 2003). [2] The construction of the HALY is illustrated in the examples in figure 1.1. For two populations with a different number of survivors and different prevalence of illness across the lifecycle, we can compute the average healthy life expectancy for each age interval.[3] By convention, comparisons of HALY are typically expressed in terms of life expectancy at birth.

---

[2] The HALY is constructed by first collecting the actual number of deaths as well as total population by each age interval. This data is used to compute the conditional probability of dying at a given age, and is then applied to an artificial cohort population (100,000 is the usual case), in order to calculate the number of persons that can be expected to survive to age x l_x. These calculations are then used to estimate the average expectation of life as $e_x = T_x/l_x$,where l_x is the number of persons surviving to age x, ${}_nL_i$ is the total number of person-years lived by persons who were alive at the beginning of the age interval x and survive to period $x+n$, and $T_x = \sum {}_nL_i, i = x, x+n, ...$,maximum life expectancy, that is, T_x is simply the total number of all future person-years lived by individuals surviving at age x. This measure of average life expectancy is then adjusted to reflect the experience of ill health: $e_x^* = \frac{1}{l_x}\sum_{i=x}^{w}(1 - {}_n\pi_i){}_nL_i$,where ${}_n\pi_i$ is the age-specific rate of being unhealthy, obtained by evaluating different health states prevalence from health surveys or clinical observations.

[3] Technical detail is in Molla *et al*, 2003.

Table 4 Existing approaches of measuring population health

CATEGORY	DESCRIPTION	TECHNIQUE	EXAMPLES	DATA SOURCES	CRITERIA SATISFIED
Health Expectancy Measures	These indices calculate an average for expected individual mortality (life expectancy) and combine this summary measure with some consideration of various dimensions of morbidity.	1.The average life expectancy for each age group is calculated. 2.For each individual, the proportion of years of healthy life to total life expectancy is computed, and the average is computed over the age cohort. 3.The results of Step1 and Step2 are combined as an arithmatic product and summed to determine the average healthy life expectancy of the population. Different age cohorts may be given different weights in summation; there is no generally-accepted approach for determining the cohort weights.	Health adjusted life years(HALY), disease adjusted life expectancy(DALE), quality adjusted life years(QALY), etc.	National age-specific mortality data (life tables), health survey data.	1, 2, 3
Healthy life gap	A family of indices that adjust measures of aggregate life expectancy to reflect the increased mortality and reduced quality of life due to the gap between perfect health and actual health.	1.Define the normative target for individual health ('perfect health'), and compute the full survivor function, assuming that all members of the population have 'perfect health'. 2.Using the life tables, compute the aggregate health gap - as measured by increased mortality and reduced functionality - due to existence of health problems. 3. Average life expectancy is reduced by the average health gap to obtain a summary measure of population health.	Health life years (HeaLY), disability adjusted life years (DALY).	Actual number of deaths for a given year in each age group, global burden of disease data	1, 2, 3
Generic Health Measurement	A family of indices which assign a cardinal measure to individual utility from health and measure the health of the population as the average of the health of individuals.	1.Calculate a cardinal measure for each individual's health, possibly reflecting multi-attribute characteristics of health. 2. Population health is the average of the health of individuals.	EuroQol (EQ-5D), Nottingham Health Profile (NHP),Rosser Index Short Form 36 (SF-36), Short Form 12 (SF-12), Sickness Impact Profile (SIP), Health Utilities Index etc.	National health survey (under a health attribute classification scheme), self-assessed health data.	1, 2, 3, 4

Population A

Age interval	Number of survivors	Rate of being healthy	Number of healthy years lived in the age interval	HALY
0--10	100	0.6	600	41.75
11--20	90	0.7	630	39.7
21--30	85	0.8	680	34.6
31--40	80	0.9	720	28.3
41--50	75	0.9	675	20.6
51--60	65	0.8	520	13.4
61--70	50	0.7	350	7

Population B

Age interval	Number of survivors	Rate of being healthy	Number of healthy years lived in the age interval	HALY
0--10	100	0.8	800	43
11--20	95	0.8	760	36.8
21--30	90	0.7	630	30.4
31--40	80	0.9	720	26.4
41--50	75	0.8	600	18.5
51--60	65	0.8	520	12.2
61--70	45	0.6	270	6

Target

Age interval	Number of survivors	Rate of being healthy	Number of healthy years lived in the age interval	HALY
0--10	100	1	1000	70
11--20	100	1	1000	60
21--30	100	1	1000	50
31--40	100	1	1000	40
41--50	100	1	1000	30
51--60	100	1	1000	20
61--70	100	1	1000	10

Fig. 1.1 Comparison of the HALYs of two populations

It is now straightforward to evaluate whether the HALY measure satisfies the criteria established in the previous section. Our first criterion is that the population health measure provide a complete ranking of all possible population health states. Since the HALY generates a cardinal measure for every possible health state for any population, it is always possible to compare the HALY of population A with the HALY of population B, and determine which is greater. Consequently, the HALY provides a complete ranking, and therefore satisfies our first criterion.

Similarly, it is straightforward to observe that criterion 2 is satisfied. Suppose that an individual in a specific age interval increases the proportion of time that s/he is healthy. Then the total number of healthy years lived in that age interval also increases, and consequently the HALY rises.

Turning to criterion 3, the methodology used to construct the adjustment to life expectancy to reflect the experience of ill health does not in fact take any account of possible interactions in health attributes. The contingent valuation methodology asks respondents to evaluate how many years of life they would forego in order to "live without arthritis" or "live without deafness", but then treats these conditions separately. In other words, the survey data does not include information on how many years of life respondents would give up in order to "live without arthritis, given that they are deaf". Therefore the HALY fails to meet criterion 3.

Likewise, criterion 4 is not satisfied. The HALY reflects only the average experience of years of healthy living in each age group; two populations with the same average number of years of healthy living in each age cohort will score the same HALY even if the distribution of those years of healthy living across members of the cohort differ significantly. For the same reason, the HALY does not meet criterion 5: mean-preserving transfers in the distribution of health across members of a given age cohort have no impact on the measured level of health of the population.

Arguably the most influential member of our second family of indices is the DALY (disability adjusted life years), [4]which resembles the HALY in several important respects.

[5]

---

[4] See http://www.who.int/healthinfo/boddaly/en/index.html

[5] The $DALY$ is measured as the sum of the years of life lost due to premature mortality in the population (YLL) plus the years lost due to disability (YLD). $YLL = N \times L$, where N is the number of deaths and L is the standard life expectancy at the age at which death occurs. $YLD = I \times DW \times L$, where I is the number of incident cases, L is the average duration of the case until remission or death (years), and DW is a weighting factor that reflects the severity of the disease on a scale from 0 (perfect health) to 1 (dead). The DALY uses the same life expectancy "ideal" standard for all population subgroups and it excludes all non-health characteristics (such as race, socioeconomic status or occupation) apart from age and sex from consideration in calculating lost years of healthy life. Similarly, it uses the same "disability weight" for everyone living a year

Population A

Age interval	Number of deaths	Life expectancy gap	YLL	Cases of disease	Disability weight	Duration (years)	YLD	DALY
0--10	10	60	600	100	0.4	5	200	800
11--20	5	50	250	45	0.3	4	54	304
21--30	5	40	200	80	0.2	6	96	296
31--40	5	30	150	60	0.1	8	48	198
41--50	10	20	200	55	0.1	4	22	222
51--60	15	10	150	70	0.4	7	196	346
61--70	50	0	0	50	0.6	7	210	210
								2376

Population B

Age interval	Numbe r of deaths	Life expectancy gap	YLL	Cases of disease	Disability weight	Duration (years)	YLD	DALY
0--10	8	60	480	90	0.3	4	108	588
11--20	7	50	350	60	0.35	3	63	413
21--30	6	40	240	70	0.4	5	140	380
31--40	4	30	120	65	0.2	6	78	198
41--50	12	20	240	50	0.1	6	30	270
51--60	20	10	200	80	0.3	6	144	344
61--70	50	0	0	40	0.5	6	120	120
								2313

Fig. 1.2 Comparison of DALYs of two populations

It is worthwhile to submit the DALY to the same scrutiny as we have imposed on the HALY. As shown in figure1.2, a DALY measure can be computed for any given population, and therefore the DALY generates a complete ranking, which means that criterion 1 is satisfied. Similarly, any improvement in the health of a single individual increases the average health of the cohort, and therefore decreases the DALY, meaning that criterion 2 is satisfied as well. As with the HALY, the conventional implementation of the DALY does not meet criterion 3, because the disability adjustment is determined separately for each possible disability, and possible interaction effects are therefore overlooked. However, more recently the morbidity adjustment used to determine the DALY has been calculated in conjunction with the Health Utility Index (discussed below), and when this is done then

---

in a specified health state.

the subsequent DALY will in fact be sensitive to changes in the combination of health conditions experienced by individuals, and thus would meet criterion 3. However, as with the HALY, the DALY fails to meet criterion 4, because it relies on averages, and therefore populations with the same average number of DALYs at every age are found to be equally healthy, even if the distribution of those DALYs changes within the population. This also means that, in general, the DALY does not satisfy criterion 5.

Finally we turn to the HUI3, which is a commonly-used member of the third family of health utilities indices (Feeny *et al*, 2002). [6] Figure 1.3 and figure 1.4 illustrate the application of the HUI3 methodology, which correspond respectively to example 1 and example 2B above in section 2, where in figure 1.3 each individual's health measure is availabe, and in figure 1.4, though it stands for the multi-dimensional case, for simplicity, only two health attributes are considered.

---

[6] HUI3 generates a cardinal measure of the well-being associated with an individual's state of health and is obtained from survey data which measure individual health with respect to eight attributes (vision, hearing, speech, ambulation, dexterity, emotion, cognition, pain), and generate utility scores (ranging from 0.00 to 1.00) for each possible health state that is, for each possible combination of these health attributes. A specific functional form (the parameters of which are estimated using the survey data) is used to obtain a cardinal measure of each individual's utility based on their observed attributes:

$$u(x) = (1/k)[\prod_{j=1}^{n}(1 + kk_j u_j(x_j)) - 1], \tag{1.1}$$

where $u(x)$ is the utility for health state x, $u_j(x_j)$ is the single attribute utility function for attribute j, k_j is the weight of attribute j, representing its importance in determining the utility score for the overall health state, k captures the interactions among attributes, $(1 + k) = \prod_{j=1}^{n}(1 + kk_j)$, if $\sum k_j > 1$,then $-1 < k < 0$,if $\sum k_j = 1$,then $k = 0$,if $\sum k_j < 1$,then $k > 0$.

	Population A U	Population B U	Population A G	Population B G
1	0.3	0.5	0.7	0.5
2	0.4	0.5	0.6	0.5
3	0.5	0.5	0.5	0.5
4	0.5	0.5	0.5	0.5
5	0.5	0.5	0.5	0.5
6	0.6	0.5	0.4	0.5
7	0.7	0.5	0.3	0.5
Total			3.5	3.5
Average			0.5	0.5

Fig. 1.3 Comparison of the single dimension HUI of two populations

K1=0.6 K2=0.7 K= - 0.71

Population A

	U1	U2	U	UA
1	0.8	0.2	0.507147	0.507
2	0.8	0.2	0.507147	
3	0.8	0.2	0.507147	
4	0.8	0.2	0.507147	
5	0.8	0.2	0.507147	
6	0.8	0.2	0.507147	
7	0.8	0.2	0.507147	

Population B

	U1	U2	U	UB
1	0.7	0.3	0.495328646	0.507
2	0.6	0.2	0.412044083	
3	0.7	0.5	0.566795076	
4	0.5	0.4	0.451750472	
5	0.8	0.3	0.538932166	
6	0.5	0.7	0.582637076	
7	0.4	0.6	0.507248166	

Fig. 1.4 Comparison of the multi-dimension HUI of two populations

Once again, it is straightforward to verify that HUI3 generates a complete ranking of the health of different populations, and that it is sensitive to the improvement in the health of any one individual, and so criteria 1 and 2 are satisfied. Turning to criterion 3, it is

evident from inspection of (1.1) that this requirement is met because the HUI3 changes as the combination of health conditions changes for the individual. However, since the HUI3 equates the health of the population with the average HUI3 across individuals, this measure obviously fails to meet criterion 4: as are shown in the figure 1.3 and figure 1.4,we see that although health is distributed differently in populations A and B, the mean health in each health attribute is equal in the two populations, and hence the use of the HUI3 leads to the conclusion that the two populations are equally healthy, which casual inspection suggests is not true. The insensitivity of the HUI3 to changes in the distribution of health across individuals also means that the HUI3 fails to meet criterion 5, because mean-preserving changes in the distribution of health across individuals does not change average HUI3.

Awareness amongst researchers of the lack of sensitivity of standard measures of population health to considerations of equality in the distribution of health across citizens underlies a recent literature on the measurement of health inequality. Wagstaff *et al.* (1991) undertake a critical appraisal of six widely accepted methods of measuring health inequality, all of which find their roots in the economic analysis of income inequality. These meaures are the range, the Gini coefficient (and associated Lorenz curve), the Pseudo Lorenz curve, the index of dissimilarity, the slope and relative indices of inequality, and the health concentration index. More recent literature includes Gakidou *et al.*(2000), which proposes using indices reflecting individual mean differences and inter-individual difference to measure health inequality. Pearcy & Keppel (2002) use the Index of Disparity, a modified coefficient of variation, to measure health disparity across populations. Pradhan *et al.* (2003) draw on the Theil entropy index to measure inequality of group health. What

these different measures enable researchers to do is to compare the extent of health inequality in different populations, which is clearly valuable. However, knowing that population A has a more equal distribution of health than population B is not enough to determine that population A is healthier than population B if the average level of health in population B is higher than in population A. In other words, what is needed is a summary measure that appropriately combines information about both the health level and the extent of health inequality. Addressing this challenge is the focus of the next section.

1.5 Alternative Approaches

The analysis of the preceeding section has shown that existing measures of population health all fail in some way or another to meet the criteria established in Section 1.3. In particular, the literature on health inequality does not provide a meaningful ranking of the health of different populations, as these measures merely rank the level of health inequality across populations. In this section, we propose two different approaches to measuring population health which satisfy the criteria laid out in Section 1.2. The first approach, which is significantly simpler to apply, draws on the literature on the measurement of poverty, and requires that there exist a cardinally-meaningful measure of the state of health of any given individual. The second approach, which exploits the techniques of stochastic dominance, allows us to compare the health of different populations even when a meaningful cardinal measure of individual health states is not available.

1.5.1 A distribution-sensitive Mean and Inequality Measure (MIM) of population health

A simple approach to integrate the measurement of health inequality into the measurement of population health is to construct a measure which expresses population health as a convex combination of mean individuals' health gap and an indicator of health inequality. This strategy is similar to that developed by Sen (1976) for the measurement of the depth and distribution (as opposed to merely the incidence) of poverty.

Assume that individual's health is measured in a way that makes the comparison possible, here for illustration, we use HUI3 system. Recall that perfect health is denoted by $\overline{x}$ and the health of a given individual by x_i. Then, using the health utility index methodology to capture the well-being associated with a specific state of health, the utility that a given individual would derive from improving their state of health from x_i to $\overline{x}$ can be expressed as $HUI3(\overline{x}) - HUI3(x_i)$. This difference is the health gap for the individual. We now define the average health gap, $\overline{G}$, as being equal to $\frac{\sum_{i=1}^{n}(HUI3(\overline{x})-HUI3(x_i))}{n}$. A distribution-sensitive measure of population health, $\widehat{H}$, can now be defined as a simple form:

$$\widehat{H} = \overline{G} + (\overline{H})I$$

where $\overline{H}$ denotes the average health utility level, and $\overline{H} + \overline{G} = H^*$, H^* is the perfect health utility. I belongs to a family of coefficients that reflects the inequality information in the health distribution among the individualsis. I may be the Gini coefficient or any of the other traditional measures of inequality like the coefficient of variation, Atkinson's index,

etc. Note that, with any of these measures of health inequality, a lower value for I implies greater equality. Consequently, a higher value for $\widehat{H}$ implies lower population health.

It is relatively straightforward to check that this measure satisfies the five criteria established above. First of all, it provides a cardinal measure of population health for every possible specification of the health of individuals, regardless of how health is distributed with respect to any health attribute. Since a number is generated for all possible states, this measure provides a complete ranking, and so criterion 1 is satisfied. Notice also that since an increase in the health of any one individual will reduce the health gap for that individual, criterion 2 is also satisfied (although it is worth recalling that $\widehat{H}$ falls when any individual's health improves). Observe also that since $\widehat{H}$ is a continuous function of $HUI3(x_i)$ for all i, and $HUI3(x_i)$ varies depending on the particular combination of individual attributes, criterion 3 is automatically satisfied. Note that the inclusion of I in the determination of $\widehat{H}$ means that the measure is sensitive to differences in the distribution of health across individuals: a more equal distribution of health leads to a lower overall score (that is, a higher level of population health). Finally, criterion 5 is satisfied as long as I satisfies this requirement; criterion 5 will not be satisfied if a measure of inequality such as the range is used, but will be satisfied if a more sophisticated measure - such as the Atkinson index - is employed.

Whereas this approach is straightforward to operationalize, it is challenging to explore its welfare ground. Although $\widehat{H}$ balances the mean and the inequality of health,(decrease in inequality can somewhat compensate the average health gap, this compensation effect is smaller when the mean health is fairly low, whereas in the case when the mean health is

high the compensation effect of the decrease in inequality would become larger, it is consistent to the observation that a healthier society would emphasize more on the actual health distriubtion, while a less healthy society is anxious to level up its average health), it is not a dicisive form that can provide an all agreeable measure of the extent of the compensation ratio.

Naturally, to overcome this weakness, it is possible to generalize it to a complex form of a distribution-sensitive measure of population health:

$$\widetilde{H} = [\alpha \overline{G}^{\rho} + \beta(\overline{H}I)^{\rho}]^{1/\rho}$$

where $\overline{G}$, $\overline{H}$ and I have the same definitions as those in the simple form, And ρ is a compensation factor that is related to the elasticity of substitution between $\overline{G}$ and $\overline{H}I$, reflecting the sensitivity of the measure to the change of health distribution. We use σ to capture the elasticity of substitution, and hence $\sigma = \frac{1}{1-\rho}$. ρ is a parameter reflecting the preference over health inequality as tradeoff with average health, and $\rho \leq 1$. As the average health level increases, more concern will be given to the inequality in the health distribution. (Bleichrodt & van Doorslaer, 2006) α and β are constants that are related to the adopted set of value judgment for the comparison. When $\rho = 1$, and $\alpha = \beta = 1$, we have the simple form of MIM. We also need to mention the fact that the compensation effect ρ in the MIM may affect the ranking over health distributions because as the associated elasticity of substitution σ ranges from infinity to 0 , the corresponding MIM functional form changes from linear form to the Leontief form, and thus exhibits Utilitarian to Rawlsian point of views over measures of average health and health inequality. Therefore, as ρ changes from negative infinity to 1 the impact of the average health gap and health inequality move in

opposite directions: i.e $\Delta\overline{G}* \Delta\overline{H}I < 0$. It is thus important to know the range of ρ for which the resulting health ranking is stable – which is clearly related to the health level, the average health difference in the population, and the relative inequality across the health distributions.

One might also suggest the application of the Foster, Greer & Thorbecker (1984) (the FGT) index for measuring population health, wherein $P_\alpha = \frac{1}{n}\sum_{i=1}^{q}(\frac{z-y_i}{z})^\alpha$, with the non-normalized version being: $P_\alpha = \frac{1}{n}\sum_{i=1}^{q}(z-y_i)^\alpha$. P_α is the level of the population health gap, n is the population size, q is the number of persons who fall below a certain health poverty line, denoted by z. Technically it is difficult to define a universally acceptable health-poverty line because of the multi-dimensional nature of health. A health-poverty line needs to identify whether the various dimensions are substitutes, complements or independent of each other. How the various elements in the health-poverty line interact has profound implications in terms of the conclusions that are reached regarding the health of any given population. From a practical point of view, z could be defined as perfect health, y_i as individual i's health level, and thus an individual's health could be defined in terms of its distance to perfect health. The FGT measure becomes a head count ratio if $\alpha = 0$,and a poverty gap if $\alpha = 1$, and with $\alpha > 1$ a less healthy individual gets a higher weight than a healthier individual.

It is quite straightforward to show that when $\alpha \leq 1$, the FGT measure is not sensitive to the distribution of health across individuals, hence it does not satisfy criteria 3, 4 and 5. When $\alpha > 1$, the FGT measure satisfies criteria 1 and 2 because it can generate a complete ranking and is pareto sensitive. If the FGT index were to use a HUI to represent

individual's health, it would directly satisfy criterion 3. Since the FGT measure is sensitive to the distribution of health across individuals, it satisfies criterion 4. Although there is not a direct component in the FGT measure to capture inequality in the health distribution, a health inequality cost could be defined as the equally distributed equivalent (EDE) health gap minus the average health gap. Therefore implicitly, the FGT measure incorporates the notion that the decrease in the health inequality cost can compensate the lost in average health. However, the tradeoff relationship between average health and health inequality is not explicitly specified in FGT. Consequently, FGT is not able to satisfy criterion 5.

Additionally, unlike income whereby an extremely poor individual could become quickly better off with an income transfer, in the health context, this cannot always happen. The existence of incurable health conditions, for instance, may thwart the useful redistribution of health resources for the current population. With the FGT approach, the weight assigned to an individual is positively associated with the individual's health gap: in some extreme instances, this may over-emphasize the importance of the least healthy individual's wellbeing in the index.

Operationalizing the MIM

To demonstrate the impact of integrating distribution considerations in measuring the health of the population, we apply the simple form of MIM measure to the examples considered earlier. The first example treated is one where two populations are considered to be equally healthy when the $HUI3$ measure is employed. As shown in figure 1.5, when

a MIM measure is used, however, population B has a lower measure of health than does population A.

	Population A	Population B	Population A	Population B	Population A	Population B
	U	U	G	G	$\bar{G}+(1-\bar{G})*gini$	$\bar{G}+(1-\bar{G})*gini$
1	0.3	0.5	0.7	0.5	0.5+(1-0.5)*0.13	0.5+(1-0.5)*0
2	0.4	0.5	0.6	0.5	0.565	0.5
3	0.5	0.5	0.5	0.5		
4	0.5	0.5	0.5	0.5		
5	0.5	0.5	0.5	0.5		
6	0.6	0.5	0.4	0.5		
7	0.7	0.5	0.3	0.5		
Total			3.5	3.5		
Average			0.5	0.5		

Fig. 1.5 Comparison of the health of two popualtions using MIM

Our second example, which is based on Table 2B and is also discussed in the context of the analysis of $HUI3$, as shown in figure 1.6, compares two populations which have the same average level of $HUI3$, although the average value for each attribute differs. Notice that the average level of the first attribute is higher in population A than in population B, whereas the reverse is true for the second attribute. But whereas $HUI3$ cannot distinguish between these two populations, the MIM concludes that population A is markedly healthier than population B.

Although these examples demonstrate that the use of a MIM will lead to different rankings for populations that a measure such as the $HUI3$ consider to be equivalent, it is evident that these rankings depend upon the particular inequality index that is chosen. If we were to use a measure such as Theil's entropy index, or the Atkinson index, in place of the Gini coefficient, the relative rankings of the populations might change in some extreme

case. Unfortunately, the theory of income inequality does not provide strong guidance with respect to identifying the "right" measure of inequality.

K1=0.6 K2=0.7 K= - 0.71

Population A

	U1	U2	U	UA	$\overline{G}+(1-\overline{G})*gini$
1	0. 8	0. 2	0. 507147	0. 507	0.493+0.507*0
2	0. 8	0. 2	0. 507147		0. 493
3	0. 8	0. 2	0. 507147		
4	0. 8	0. 2	0. 507147		
5	0. 8	0. 2	0. 507147		
6	0. 8	0. 2	0. 507147		
7	0. 8	0. 2	0. 507147		

Population B

	U1	U2	U	UB	$\overline{G}+(1-\overline{G})*gini$
1	0. 7	0. 3	0. 495328646	0. 507	0.493+0.507*0.1
2	0. 6	0. 2	0. 412044083		0.5437
3	0. 7	0. 5	0. 566795076		
4	0. 5	0. 4	0. 451750472		
5	0. 8	0. 3	0. 538932166		
6	0. 5	0. 7	0. 582637076		
7	0. 4	0. 6	0. 507248166		

Fig. 1.6 Comparison of the health of two populations using MIM

1.5.2 Stochastic-dominance and the measurement of population health

As discussed above, the main weakness with the MIM approach to measuring population health is that reasonable people may disagree about whether or not health inequalities are best captured by the Gini coefficient versus Theil's entropy measure, etc, and thus to the extent that these differences lead to disagreement about the rankings of the health of different populations, it can be difficult to reach agreement about important matters of policy. In the literature on poverty and income inequality, a possible solution to this impasse is in

the development and application of the technique of stochastic dominance analysis. An additional benefit of using this technique is that it is possible to compare multi-dimensional distributions even in the absence of a satisfactory cardinalization of the utility associated with an individual's state of health.

It is useful to provide an intuitive explanation of the conceptual underpinnings of stochastic dominance analysis. For convenience, let us consider two populations with an equal number of individuals, ranked from sickest to healthiest. The simplest application (stochastic dominance of order 1) counts the number of individuals whose level of health is lower than some reference level in population A and compares that to the number of individuals with a level of health less than this same reference level in population B[7]; this comparison is carried out for a range of possible reference levels (for example, from zero to perfect health). If fewer individuals in population A always have health below the reference level than in population B, then population A stochastically dominates population B. When we consider stochastic dominance of order 2, we consider the cumulative distribution of the gaps between the level of health of each individual and successive reference levels of health. The population with smaller gaps is judged to be healthier. If the two distributions intersect, we can extend this approach to carry out higher-order comparisons.

Applying stochastic dominance techniques when individual well-being can be cardinalized

We can use stochastic dominance to compare two distributions (A, B) of population health, when we are able to measure the well-being of a given individual in a cardinal

[7] Note that the technique is also applicable to two populations that are not having the same number of individuals. In that case, it is the "proportion" instead of the "number" of individuals that are being counted.

fashion like using the $HUI3$ index. Define the cumulative distributions C_A and C_B. Let $D_A^1(x) = C_A(x)$, that is, $D_A^1(x)$ is the proportion of population A with a $HUI3$ score less than or equal to x. Define the s^{th} order health index $D_A^s(x) = \int_0^x D_A^{(s-1)}(t)dt$ for any integers $s \geq 2$. As proved by Fishburn(1976), $D_A^s(x) = \int_0^x D_A^{(s-1)}(t)dt$ is equivalent to $D_A^s(x) = \frac{1}{(s-1)!}\int_0^x (x-t)^{s-1}dC_A(t)$. And $D_B^s(x)$ is defined analogously.

Definition 1 *Distribution C_B stochastically dominates distribution C_A at order s if $D_A^s(x) \geq D_B^s(x)$ for all x.*

It is very important to examine whether stochastic dominance satisfy the five criteria previously set out to evaluate the approaches of measuring population health. As defined, when the individual's health is measured by the $HUI3$, the health of the population is a uni-dimensional construct. Since stochastic dominance techniques can be used to do pair-wise comparisons of health distributions, it generates a complete ranking over all possible health distributions in the comparison order s. Consequently, criterion 1 is satisfied. Criterion 2 is also met due to the fact that any increase in the individual's health moves the distribution curve to the right of the original one, and hence brings a higher ranking in terms of stochastic dominance. Moreover, since we are using $HUI3$ to aggregate the health attributes together into a single measure of individual health, and it is the $HUI3$ scores that are used to derive the health distribution curves, so stochastic dominance automatically meets criterion 3. Of course, by construction, stochastic dominance is distribution sensitive, hence criterion 4 is also satisfied. Finally, with respect to criterion 5, if there is a mean-preserving change towards the mean in the population health distribution, this creates

a "dent" in the lower end and a "hump" in the upper end of the original population health distribution; when checking higher order stochastic dominance, the curve corresponding to the new distribution naturally lies below the original one, demonstrating that the population is healthier, and so the Pigou-Dalton principle is satisfied. And it still meets criterion 5 even when the change in distribution decrease the mean while decreasing the inequality to a certain extent, i.e pro poor health individual.

Operationalizing stochastic dominance techniques when individual well being can be cardinalized

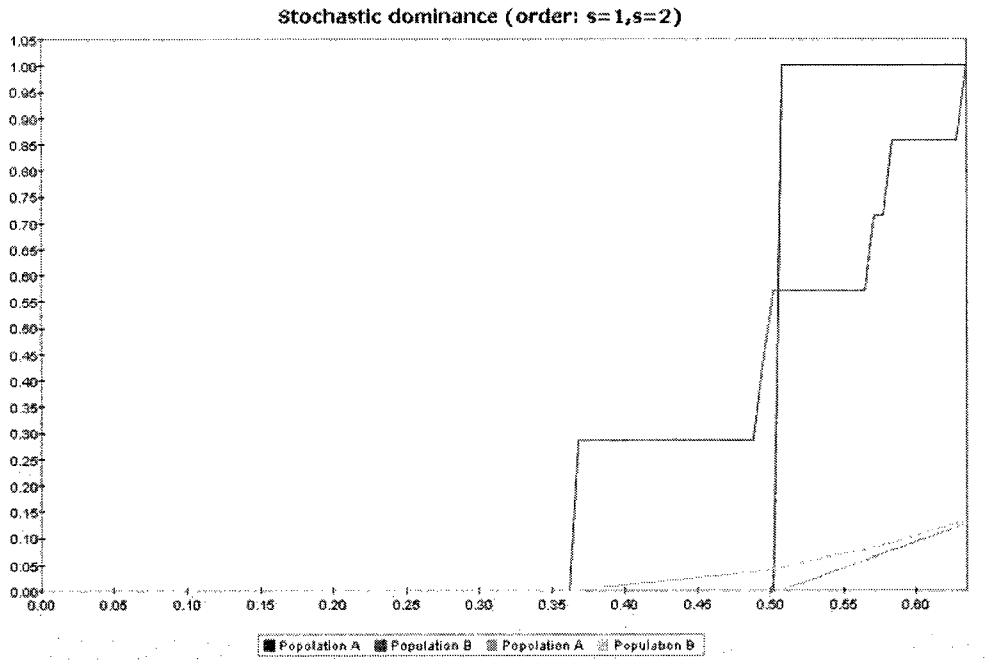

Fig. 1.7 Comparison of the health of two populations using stochastic dominance

The application of stochastic dominance relationships to comparing the health of populations can be illustrated by considering Example 2B. As illustrated by figure 1.7, when considering only first-order stochastic dominance, the two cumulative health distrib-

ution curves intersect each other. Going on to consider second order stochastic dominance, the curve for population A lies below that for population B. Consequently, population A stochastically dominates population B, which means that population A is healthier than population B.

Additionally, even in the unidimensional context, there are other situations when the application of stochastic dominance provides additional insights into the comparison of the health of different populations. Consider, for example, figure 1.8, which describes two populations with the same mean level of health as measured by $HUI3$.

Population A

	U(P)	U(R)	U(S)	U(H)	U(P,R,S,H)
1	0.1	0.2	0	0.3	0.3131962
2	0.1	0.2	0.5	0.5	0.5090672
3	0.1	0.2	0.5	0.5	0.5090672
4	0.38	0.32	0.25	0.5	0.6148886
5	0.66	0.44	0	0	0.6483448
6	0.66	0.44	0	0	0.6483448
7	0.66	0.44	0.5	0.3	0.7620486
Average	0.38	0.32	0.25	0.3	0.57

Population B

	U(P)	U(R)	U(S)	U(H)	U(P,R,S,H)
1	0.1	0.44	0	0.3	0.4552943
2	0.1	0.44	0.15	0.3	0.4887962
3	0.38	0.32	0.1	0.3	0.5400167
4	0.38	0.32	0.1	0.3	0.5400167
5	0.38	0.32	0.25	0.3	0.569741
6	0.66	0.2	0.55	0.3	0.7037437
7	0.66	0.2	0.6	0.3	0.7118131
Average	0.38	0.32	0.25	0.3	0.57

Fig. 1.8 Comparison of the health of two populations

A comparison of the two populations using first-order stochastic dominance yields figure 1.9, with the darker line representing the CDF of the health of population B.

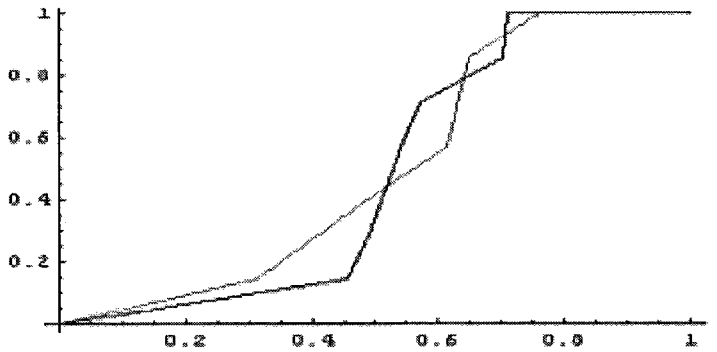

Fig. 1.9 CDFs of the health distributions of two populations

The application of stochastic dominance of order 1 reveals that there are several intersections for the two cumulative density functions; consequently, we consider second-order (and, if necessary, third-order, etc.) stochastic dominance until there is no intersection of the curves along the support.

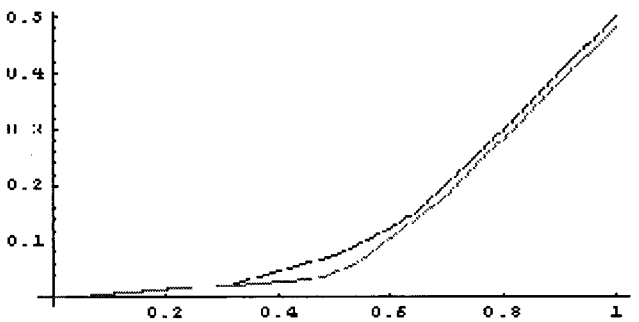

Fig. 1.10 Stochastic dominance curves(order=2) of the health distributions of two populations

Figure 1.10 illlustrates the relationship between the distributions when viewed at the level of second-order stochastic dominance. The cumulative health curve of population B lies everywhere below that of population A for the whole support, hence leading to the conclusion that the distribution of health of population B stochastically dominates population A: population B is healthier than population A.

Applying stochastic dominance techniques with multidimensional health states

As noted in our discussion above, it is not always possible to derive a cardinally-meaningful measure of multidimensional individual health states. For example, although the Health Utility Index is based on survey data, there remains some substantive uncertainty as to the meaningfulness of the cardinalization derived from the application of the HUI measure of individual's state of well-being. Consequently, it is useful to develop an approach to the measurement of population health which does not rely upon the existence of a cardinal summary measure of individual health states. A multi-dimensional version of stochastic dominance approach might be able to build up a distribution-sensitive measure of population health when a cardinal measure of individual well-being is not available.

Duclos *et al* (2006) extend the stochastic dominance approach to the multidimensional case, and illustrate the dominance condition in two dimension setting. Applying their methodology to the analysis of population health, suppose H_A and H_B are the multivariate health distributions for population A and population B. Let $D_A^1(h_1, h_2, ..., h_n) = H_A(h_1, h_2, ..., h_n)$, $H_A(h_1, h_2, ..., h_n)$ is the joint cumulative health distribution of n dimensions of population A. Define $D_A^s(h_1, h_2, ..., h_n) = \int_0^{h_1} \int ... \int_0^{h_n} D_A^{(s-1)}(t_1, t_2, .., t_n) dt_1 dt_2 dt_n$.

This is equivalent to $D_A^s(h_1, h_2, ..., h_n) = \frac{1}{n(s-1)!} \int_0^{h_1} \int ... \int_0^{h_n} (h_1 - t_1)^{s-1} (h_2 - t_2)^{s-1} ... (h_n - t_n)^{s-1} dH_A(t_1, t_2, .., t_n)$. $D_B^s(h_1, h_2, ..., h_n)$ is defined analogously.

Definition 2 *If for all* $(h_1, h_2, ..., h_n) \in [\underline{h}, h^*]^n$, $D_A^s(h_1, h_2, ..., h_n) \geq D_B^s(h_1, h_2, ..., h_n)$ *in the entire support, then* H_B *stochastically dominates* H_A *at order* s.

Our five criteria are also satisfied in this extended stochastic dominance approach. For criterion 1, though it is dealing with a multidimensional space, stochastic dominance undertakes pairwise comparisons of the health of populations, so generates a complete ranking over all possible population health distributions. Similarly, it is straight forward to see that criterion 2 is met since the multidimensional extension of the stochastic dominance approach retains the monotonicity property. For criterion 3, the derivation of the multivariate distribution of population health implicitly attaches weights to the health gap on an attribute-by-attribute basis. Consequently, the combination of health attributes makes a difference in the calculation of the measure of population health. As before, criteria 4 and 5 are quite obviously satisfied. In particular, a transfer within different health attributes would change the shape of the multivariate distribution of the population health, hence brings the stochastic dominance for the distribution after change over the orginal distribution in higher order.

Notice that although this multidimensional approach allows us to compare the health of populations in more than one dimension - therefore not obliging us to rely on any particular cardinalization of individual well-being - the weakness with this approach is precisely that the weight attached to inequality in the distribution of any given attribute is driven by the actual distribution of the attributes, rather than reflecting any exogenous information about which attributes may be "more important" than others, i.e., we may "know" that vision is a more important attribute than "mobility" but depending on the actual distribu-

tion of the attributes inequality in mobility may end up being weighted more heavily than inequality in vision[8].

Operationalizing the stochastic dominance approach with multidimensional health states

The application of stochastic dominance relationships in the absence of a means to cardinalize individual health states can be illustrated by considering Example 2B, limited to the first two dimensions, and graphing the two distributions in 3 dimensional space. Figure 1.11 illustrates the relationship between the two distributions, when looking at the order-two dominance relationship.

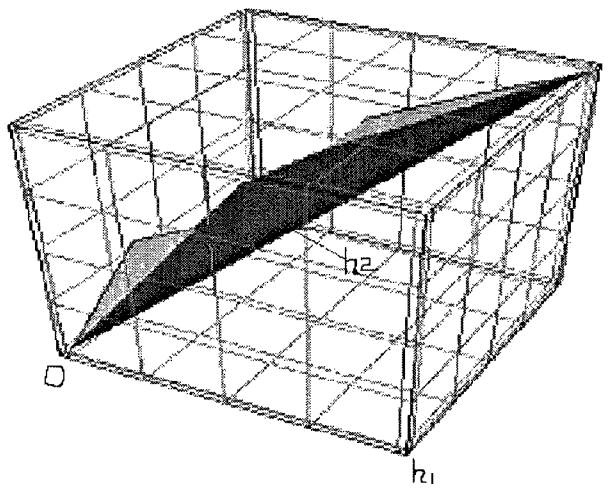

Fig. 1.11 Illustration of stochastic dominance upon two health attributes

Observe that the distribution for population A lies above that for population B up to a certain level, then below, then once again above, indicating the surfaces intersect. Consequently, to establish which population is healthier, it is necessary to consider a higher-order

[8] Developing a methodology for re-weighting each of these dimensions is an interesting direction for future research.

comparison. In contrast, if the objective is to limit the comparison of the populations to focus on the "least healthy", and the cut-off point for "least healthy" lies below the the level where the two surfaces intersect for the first time, then population A is stochastically dominated by population B, and is therefore less healthier.

1.6 Conclusion

This paper considers the problem of developing a distribution-sensitive measure of population health when individudal health states are described by many different health attributes. Whereas the existing literature on the measure of population health focuses largely on the problem of developing a reasonable cardinal measure of individual health, the problem of aggregating these measures of individual health in a way that reflects inequality in the distribution of health across individuals has received little attention. Indeed, the literature on health inequality is principally concerned with developing appropriate measures of disparity in the distribution of health across a population, rather than in trying to assess how measures of health inequality should be combined with measures of individual well-being to generate a meaningful measure of the health of a population.

This paper argues that a satisfactory measure of population health must satisfy five different criteria. Commonly-used measures of population health are evaluated according to these criteria, and are found to be wanting to differing degrees. Subsequently, we propose two different approaches to developing a distribution-sensitive measure of population health, one of which requires that a cardinal measure of individual well-being be available (such as $HUI3$), and a second - stochastic dominance - which relies only on information

about the distribution of attributes. These two approaches constitute a significant advance with respect to existing methodologies, and should prove useful in policy-making contexts.

Chapter 2
Robust Distributionally-Sensitive Comparisons of Population Health

2.1 Introduction

Building on the achievements of the literature on the measurement of poverty and of income inequality, there is an extensive literature on health inequality which uses indices such as the Gini coefficient, the coefficient of variation, Theil's index, or the health concentration index to rank different populations with respect to the extent to which health is equally distributed (Wagstaff et al, 1991; Mackenbach and Kunst, 1997; Gakidou et al, 2000; Wagstaff, 2000; Wagstaff, 2002; Clarkea et al, 2002; van Doorslaer and Jones, 2003; Regidor, 2004; Koolman, 2004). For policy makers, the plethora of alternative indices of inequality is somewhat bewildering, and it is not evident which index is 'best'. Moreover, as the income inequality literature has stressed (Kolm, 1969; Atkinson, 1970; Sen, 1973; Blackorby and Donaldson, 1978; Weymark, 1999; Bommier and Stecklov, 2002; Bleichrodt & van Doorslaer, 2006), different views of social welfare are embedded in each of these specific inequality indices. For this reason, investigations of income inequality now typically compare different distributions using stochastic dominance criteria (Shorrocks, 1983; Foster, 1985; Shorrocks and Foster, 1987; Davies and Hoy, 1995; Moyes, 1999), which are viewed as providing a more robust ranking of alternative distributions, and it is therefore natural to evaluate the suitability of this technique for comparing the health of populations.

Whereas there are obviously many conceptual similarities between the measurement of income inequality and health inequality, there are at least two significant differences. Firstly, it is meaningful to talk about being in excellent (or perfect) health, whereas it is obviously absurd to talk about having excellent (or perfect) income. Therefore, whereas measures of income inequality are inherently relativistic, it is clearly meaningful to talk about individuals being in poor health (and, in particular, individuals in very poor health end up dead). Secondly, whereas income can always be redistributed from rich to poor - albeit at a considerable cost in terms of economic efficiency - there is no sense in which health is necessarily transferrable across individuals. In particular, there are many individuals who suffer from chronic or otherwise debilitating conditions and whose health status cannot be improved, regardless of how much money is spent trying to ameliorate their situation. In contrast, the incidence of other conditions, and in particular the health of 'typical' individuals, may be directly affected by public policy decisions (e.g., to finance the provision of widespread vaccinations). It is therefore essential that the measurement of health inequality in different populations be sensitive to changes in the distribution of health amongst those individuals whose health status is indeed sensitive to policy decisions, and that possibly relatively less weight should be placed on the health of individuals who may objectively be less healthy, but whose health status is not sensitive to changes in health policy.

The analysis below seeks to contribute to the analysis of the appropriate methodology for comparing the health of populations by comparing stochastic dominance, Lorenz dominance (sequential comparison) and the Mean-Inequality Measure (MIM). In particular, we try to understand the circumstances under which these alternative approaches rank

populations similarly, and when the rankings they generate will differ. We also seek to clarify the implicit social welfare judgements embedded in these measures. We show that when populations are compared using stochastic dominance criteria, that very high weight is implicitly placed on the health status of the least-well-off individuals in the population.

Whereas the theoretical analysis clarifies the underlying differences between the stochastic dominance approach and the MIM, it cannot be assumed that these differences are empirically significant. Unless these measures rank populations differently in actual real-world settings, the choice of stochastic dominance over MIM (or of MIM over stochastic dominance) is essentially a matter of taste, or of operational convenience. Consequently, in section 4, we illustrate the differences between these two approaches using Canadian and US data. This exercise demonstrates that these two approaches will typically generate different rankings, in particular because the rankings generated using stochastic dominance are largely determined by the number of individuals in very poor health in each population under consideration. This fact, combined with the ethical transparency of the MIM measure, lead us to conclude that the MIM may be more appropriate than the application of stochastic dominance criteria when comparing the health of populations.

2.2 Dominance Conditions in Comparing the Health of Populations

Our objective in this section is to compare and contrast Lorenz dominance and stochastic dominance for the purpose of comparing the health of populations, drawing attention to the implicit view of economic welfare that underlies each of these measures. Atkinson

(1970) first applied the approach of relative Lorenz dominance to the analysis of income inequality, stimulating enormous interest in the problem of ranking distributions for the purpose of studying income inequality and poverty. Excellent surveys can be found in Kakwani (1980) and Lambert (1989). Relative Lorenz dominance was relatively quickly shown to generate only partial orderings, and therefore significant effort was invested in developing extensions of this concept so as to generate unambiguous rankings over distributions. Shorrocks (1983) developed the concept of generalized Lorenz curves and the criterion of absolute Lorenz dominance. Shorrocks and Foster (1987) imported the stochastic dominance criterion from the literature on the measurement of uncertainty, thereby greatly expanding the capacity of researchers to rank different income distributions, and also established the equivalence of second order stochastic dominance and absolute Lorenz dominance.

It was a natural step for researchers interested in health and inequality to start to look at Lorenz curves for the distribution of health in a given population. Le Grand (1989) explores the use of the Lorenz curve and the Gini coefficient to measure the inequality of health distributions, and Wagstaff et al (1991) also discusses the underlying implication and weakness in applying Lorenz curve to measure health inequality. There are, however, some conceptual issues which arise with respect to the appropriate ranking of distributions of population health which are not of equal importance when considering the ranking of different distributions of income. For example, it is (relatively) meaningful to talk about an individual being in 'perfect health', and of the gap between a given individual's actual state of health and the ideal state of 'perfect health'. In comparison, there is no sense in which

one can describe a 'perfect income', and therefore the gap between high income and low income individuals in a given community has no intrinsic significance, except insofar as it is indicative of the extent of income inequality in the community. Another key difference is that whereas problems of poverty can be attenuated if poor individuals are given more resources, the same is not true of health: many health problems are - at least at the present time - insoluble, and therefore additional health care spending on these individuals will not improve their state of health (although giving such individuals high incomes might improve their overall level of well-being). For these reasons, amongst others, it is clearly useful and important to carefully examine the appropriateness of different approaches to rank the distributions of health in populations.

2.2.1 Lorenz Dominance

In the context of the measurement of population health, there has been very little use made of the absolute Lorenz curve. Instead, researchers have relied principally on the relative Lorenz curve, which is well-suited to describing intra-population inequality in the distribution of the parameter of interest. In constrast, the absolute Lorenz curve, and the associated concept of absolute Lorenz dominance, is particularly appropriate when considering comparisons across populations, or of a given populations at different points in time, because it accords importance to the proportion of the population which meets specific performance criteria in terms of health. To construct the absolute Lorenz Curve, citizens are ranked in terms of health level, from lowest to highest. Then, for example, we can compare the average level of health of citizens belonging to the lowest quantile across different populations.

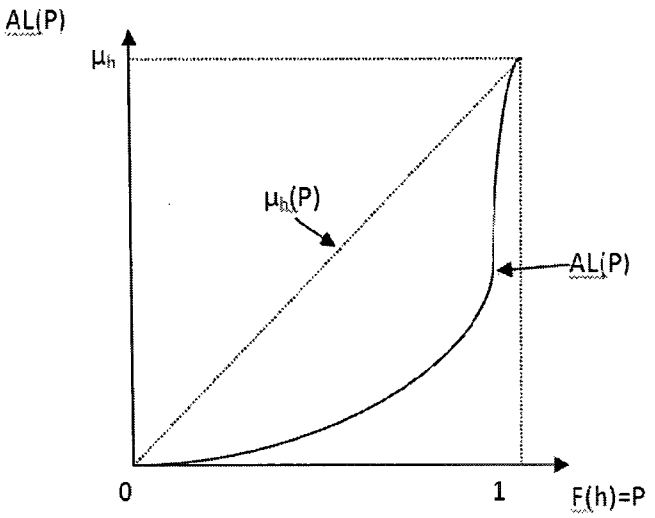

Fig. 2.1. Absolute Lorenz Curve

If the average level of health of citizens in every quantile is higher in population A than in population B, then population A is said to Lorenz-dominate population B.

Definition 3 *Assume that an individual' s health status is denoted by* $h_i, h_i \in [H_0, H^*]$, *and let* $F(h)$ *be the cumulative distribution function of the health of the population, and* $F^{-1}(P) = \{h \mid f(h) = P(t \leq h)\}, 0 \leq P \leq 1$. *Then the absolute Lorenz Index is a function* $AL(P) = \int_0^{F^{-1}(P)} hdF(h)$, *where* P *is the cumulative probability, and* $0 \leq P \leq 1$.

Consider an arbitrary proportion P of the least-healthy individuals in a given population. Then, as illustrated by Figure 2.1, $\mu_h(P)$ denotes what the average health level would be of the population if health were distributed equally across all of these individuals, that is, $\mu_h(P) = \mu_h \times P$. In constrast, we denote the actual average health level of the propor-

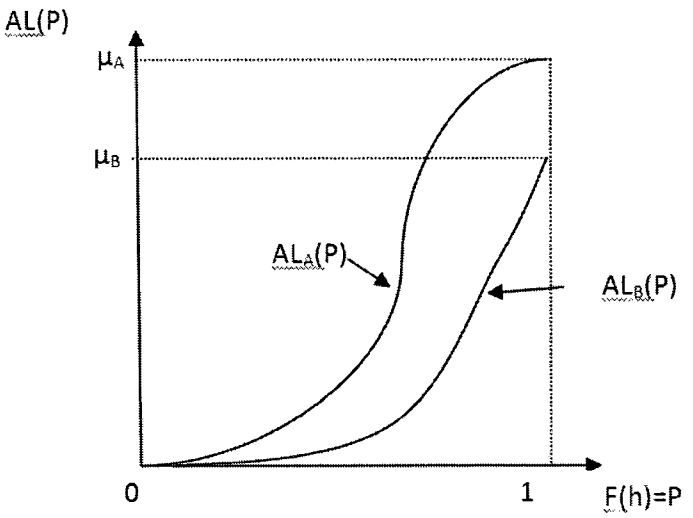

Fig. 2.2. Absolute Lorenz Dominance

tion P of the least-healthy individuals in the general population as $AL(P)$. Consequently, for any P, $AL(P) \leq \mu_h(P)$.

Definition 4 *Now consider two populations A and B, with associated health distributions $F_A(h)$ and $F_B(h)$.If $AL_A(P) \geq AL_B(P)$for all P, and $AL_A(P) > AL_B(P)$ for some P, health distribution A absolute Lorenz dominates health distribution B.*

Absolute Lorenz dominance holds when the absolute Lorenz curve of the distribution of health in population A always lies above that for population B. This means that, for any given P, the average health level of the least-healthy proportion P of the population is higher in A than in B.

When the ALC is standardized by being divided by the average health level of the whole population, this generates the relative Lorenz Curve (RLC), which is commonly

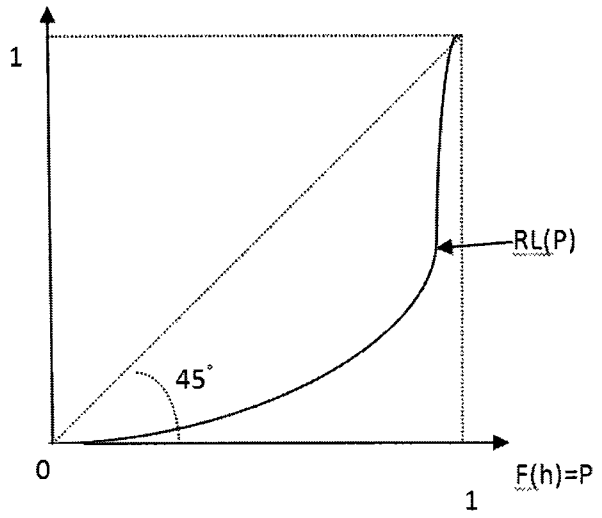

Fig. 2.3. Relative Lorenz Curve

used (Le Grand (1989), John Wildman (2003), etc.) for examining intra-population inequalities in the distribution of health. The Lorenz Index is denoted as $RL(P)$, where

$$RL(P) = \frac{1}{\mu}\int_0^{F^{-1}(P)} hdF(h)$$

and

$$\mu = \int_0^1 hdF(h).$$

In comparing the suitability of the ALC and the RLC for the purposes of comparing the health of populations, the biggest drawback to the use of the RLC is that it does not take account of the objective information regarding the absolute levels of health of citizens, and merely reflects the dispersion in the distribution of health. Therefore, consider the following two populations, A and B, where the health of individuals in population A is described by $(1, 5, 7, 9, 11)$ and in population B by $(3, 15, 21, 27, 33)$. Using the RLC, these populations are judged to be equivalent. Clearly, however, a useful approach to

ranking the health of populations needs to conclude that population B is healthier than population A.

Remark 1 *Consider two populations A and B, if $\mu_A \geq \mu_B$, and $F_A(P) \leq F_B(P)$ for all P, then $AL_A \geq AL_B$.*

Proof. See Appendix. See also the discussion in Atkinson (1970), Shorrocks (1983). ■

Whereas Remark 1 provides a computationally useful test for checking for absolute Lorenz dominance, in actual application it is generally unlikely to be of much use, principally because it is unusual for a given dataset to exhibit this non intersection property. As has been demonstrated repeatedly in the literature on income inequality, when comparing two populations the typical pattern is that the Lorenz curves cross at least once, and often several times, which means that a more sophisticated approach to comparing the distributions is required. In the literature on income inequality, this problem is often now addressed by using the criteria of stochastic dominance in order to compare two populations for which the Lorenz curves intersect. In addition, when we use the five criteria set forth in Chapter 1 to evaluate the appropriateness of ALC and RLC to measure population health, it is easy to find out that although they satisfy the last four criteria, they both violate the first criterion because they fail to generate complete ranking for the comparison of different health distributions in the case when there are multiple intersections between two Lorenz curves. In contrast, as discussed in Chapter 1, stochastic dominance does meet the criteria.

2.2.2 Stochastic dominance

In the context of health, the application of the stochastic dominance criterion involves examining the proportion of people whose health status falls below a given health level h. In comparing the distribution of health in two populations, A and B, if there is a lower percentage of the population with a health status less than any given health level h in population A as compared to population B then population A is considered to be healthier. This conclusion reflects the fact that the probability that the health status of a randomly selected individual of the population falls below h is lower in population A than it is in population B.

Definition 5 *Let $F(h)$ be the cumulative distribution function of health in a given population. Then the health dominance index is $H^1(h) = F(h)$,and $H^s(h) = \int_{h^0}^{h} H^{s-1}(t)dt$ for $s \geq 2$, where h^0 is the lowest health level.*

Here, s can be viewed as a measure of the dimensions of the distributions which we are comparing to get a better insight into the intrinsic health of the populations under consideration. The health dominance curve of order s corresponds to a series of health dominance indices which are successively integrated from order 1 to higher orders of dominance. Consequently, suppose we have two populations A and B, with health cumulative distributions denoted by $F_A(h)$ and $F_B(h)$ respectively. Using the criterion of stochastic dominance, of order 1 up to order s, we can compare the distribution of health in these populations with respect to a particular reference level of health h.

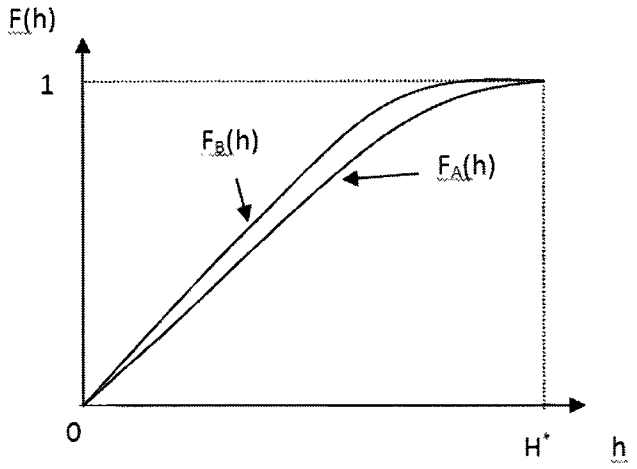

Fig. 2.4. Stochastic dominance of order 1

Definition 6 *$F_A(h)$ stochastically dominates $F_B(h)$ at order s, if $H_B^s(h) \geq H_A^s(h)$ for all h.*

In the case that the health cumulative distribution curves don't intersect, the comparison stops at order 1. Otherwise, it is necessary to consider successively higher orders of stochastic dominance, stopping the process only at the point where there is no intersection between the health dominance curves relative to this level of comparison. As is illustrated by figures 2.4 and 2.5, there is dominance of order s when the health dominance curve of order s for one population does not intersect the health dominance curve for the other population across the entire support.

For the purposes of interpretation as well as for statistical inference, it is useful to express the stochastic dominance criterion using an expression which depends upon the cumulative distribution of population health.

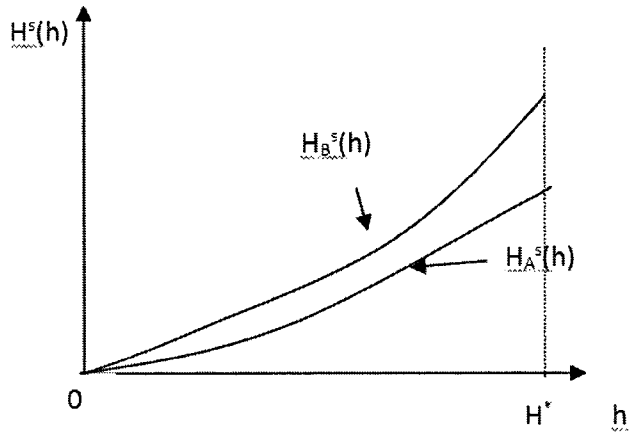

Fig. 2.5. Stochastic dominance of order s

Definition 7 *Let $F(h)$ be the cumulative distribution function of a population health. Define $\widetilde{H}^s(h) = \frac{1}{(s-1)!}\int_{h^0}^{h}(h-t)^{s-1}dF(t)$. Then $\widetilde{H}^s(h)$ is equal to $H^s(h)$ at any order.* [9] *Consequently, $F_A(h)$ stochastically dominates $F_B(h)$ at order s if $\widetilde{H}_B^s(h) \geq \widetilde{H}_A^s(h)$ for all h.*

It is easy to see that first order stochastic dominance simply implies that, with respect to any reference level of health h, the population which is characterized by the larger health gap is considered to be dominated: in the dominated population, there is always a higher proportion of the population with a level of health equal to or less than h. Similarly, when comparing two populations using the criterion of second order stochastic dominance for any reference level of health h, the criterion first ranks individuals with respect to their state of health (from lowest to highest) and then compares the average health level of the

[9] See Fishburn (1976) for the proof of their equivalence.

proportion P of the left-hand tail of the distribution of health in each of the populations. Interpretation of higher orders of stochastic dominance is relatively difficult.

2.2.3 Linking Absolute Lorenz Dominance, Stochastic Dominance, and the Measurement of Social Welfare

In this section, we present some results which, on the one hand, link together the criteria of absolute Lorenz dominance (ALD) and the social welfare functions representing the ALD criterion and, on the other hand, link together the stochastic dominance (SD) criterion and the social welfare functions which provide an equivalent ranking of the health of different populations. The usefulness of this exercise lies in the fact that it provides us with greater insight into why the ALD and SD criteria rank the health of populations in somewhat different ways.

Atkinson (1970) was the first to notice that absolute Lorenz dominance (ALD) and second order stochastic dominance (SD) generate identical rankings of different distributions, pointing out these two approaches are equivalent to each other since integrating the absolute Lorenz index by parts yields

$$\begin{aligned} AL(P) &= \int_0^{F^{-1}(P)} h dF(h) \\ &= hF(h) - \int_0^h F(h)dh \\ &= hF(h) - H^2(h) \end{aligned}$$

As shown in Figure 2.6, the ALD index at $F(H')$ is the area above the cumulative distribution curve from 0 to $F(H')$. This area is equal to the size of the square H' by $F(H')$

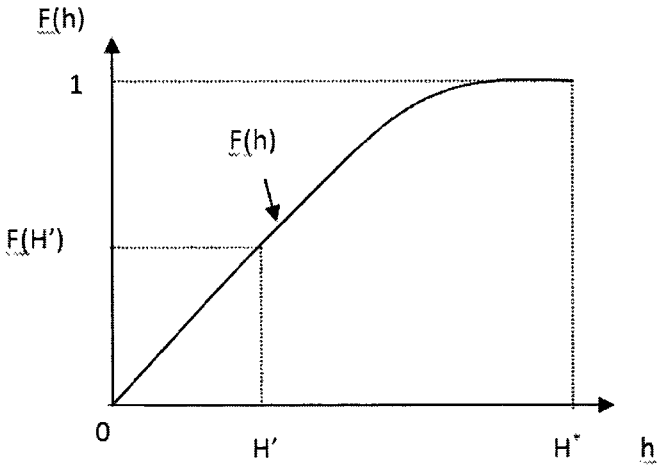

Fig. 2.6. Link between ALD and SD

minus the area below the cumulative distribution curve from the lowest health level to H', which is exactly the second order SD index at H'. So the ALD and SSD criteria necessarily rank distributions in the same way; the difference is that the dominated distribution has a smaller ALD index, but a larger SSD index. Note that the non-intersection of absolute Lorenz curves is necessary for ALD, and is therefore equivalent to the non-intersection of second-order stochastic dominance curves. Whenever the absolute Lorenz curves intersect, there is no longer ALD, nor is there SSD.

Of course, in empirical contexts, it is typical that the Lorenz curves will intersect, and therefore it is necessary to strengthen the ranking criterion. Shorrocks and Foster (1987) apply "transfer sensitivity" to Lorenz dominance, which modifies the Pigou-Dalton criterion in such a way as to accord extra weight to transfers occurring lower down in the distribution. They restrict attention to situations where the Lorenz curves cross only once. They prove that when the relative Lorenz curve for population A intersects that for population B

from above and has lower variance σ^2(when the means of both distributions are equal) or a lower coefficient of variation $\frac{\sigma}{\mu}$(when the means are not equal), then population A dominates population B for all transfer-sensitive measures of inequality. In fact, the transfer sensitivity criterion is consistent with third order stochastic dominance condition. They demonstrate that the problem of comparing distributions when the relative Lorenz curves cross only once is equivalent to requiring that third order stochastic dominance holds.[10]

However, as discussed above, in undertaking empirical comparisons of different distributions of health in populations, use of the absolute Lorenz curve is preferred to the use of the conventional Lorenz curve because the measurement of individual health levels is cardinally meaningful. And given that absolute Lorenz curves are also likely to intersect, it is useful to build on Shorrock and Foster's (1987) work in the inequality context to develop a simple test for third order stochastic dominance using the absolute Lorenz curves.

Proposition 1 *Given two health distributions $F_A(h)$ and $F_B(h)$, if the corresponding absolute Lorenz curve of A (ALC_A) intersects that of B (ALC_B) once from above, then A is preferred to B in the sense of third order dominance if and only if the difference of their variances is larger than the difference of their squared average health gaps, i.e., $\sigma_B^2-\sigma_A^2 \geq \overline{g}_A^2-\overline{g}_B^2$, where $\overline{g}_j$ denotes the average health gap in population j. Alternatively, A is preferred to B in the sense of third order dominance if and only if $\frac{\sigma_B-\sigma_A}{\overline{h}_B-\overline{h}_A} \geq \frac{\overline{g}_B+\overline{g}_A}{\sigma_B+\sigma_A}$.*

Proof. See Appendix. ■

[10] Note that only when the distribution with lower variance crosses from above can it be claimed to dominate the other distribution. If it crosses from below, the ranking of the distributions is still ambiguous.

This proposition can be compared to the result established by Foster and Shorrocks (1987) which examines the relationship between the relative Lorenz curves and third order stochastic dominance. Notice that if the mean health gap is equal in the two populations, then - as with Foster and Shorrocks (1987) - the population with the lower variance is healthier. However, in the case where the means are not equal, then the criteria for third-order stochastic dominance differ. With relative Lorenz curves, the key comparison is between the coefficient of variation in the two populations; with absolute Lorenz curves, the gap in the variances is compared to the difference in the square of the average health gap in each population. Conceptually, what this reflects is the tradeoff between a lower average health level and reduced health disparity. In the case where the average health gap in population A is greater than that of population B, but where the variance in the distribution of health in population A is smaller than that of population B, the gap in the means may be offset by the gap in variances.

The problem is, of course, that the intersection condition will generally not be satisfied, and therefore successive pairwise comparisons of Lorenz indices will typically generate at best a partial ordering of distributions of population health (Shorrocks, 1983; Zoli, 1999). Davies and Hoy (1995) push forward the theory of ranking income distributions using the relative Lorenz dominance criterion by developing a test for establishing dominance when the relative Lorenz curves intersect multiple times. They articulate a concept of aversion to downside inequality (ADI), which is analagous to the concept of transfer sensitivity. Two distributions f and g can be unanimously ranked with respect to all inequality indices that are sensitive to the transfer principle and ADI when their Lorenz curves cross at least

once (let f initially cross g from above) and their means are the same. The sufficient and necessary condition is that the variance of f over all the crossing points (variance within the subpopulation defined by $P \in [0, P_i]$, P_i denotes the series of crossing points) is less than or equal to that of g. In effect, what Davies and Hoy's (1995) result demonstrates is that when the partial means are equal for each subpopulation defined by the crossing points for the relative Lorenz curves, then if the second moment of the health gap for one population is always larger than the second moment of the other population, then this implies that the variance over this interval must also be larger.

Unfortunately, an equivalent result cannot be established for absolute Lorenz curves which intersect multiple times. The barrier to establishing a simple test for determining whether or not one distribution dominates the other is that the means of the subpopulations cannot be equal. To compare distributions which have intersecting absolute Lorenz curves requires the use of higher order stochastic dominance criteria.

For the purposes of policy-making, it is particularly useful to characterize a social welfare function that ranks alternative distributions of population health in the same way as the absolute Lorenz dominance criterion.

Remark 2 *(Shorrocks, 1983) Consider a population of n individuals, and consider an arbitrary division of this population into k sub-populations, $\{n_1, ..., n_k\}$ where each successively larger sub-population includes the smaller ones, i.e., if individual $i \in n_1$ then $i \in n_j$ for all $j = 1, ..., k$. Then there exists a sequence of utilitarian social welfare functions, $[W(\overrightarrow{h_1}), ..., W(\overrightarrow{h_n})]$, where $W(\overrightarrow{h_j}) = \sum_{i=1}^{n_j} U(h_i), j = 1 \ldots k$ which is equivalent*

to the discrete representation of the absolute Lorenz index, $AL(P) = \sum_{j=1}^{J} h_j \frac{k_j}{n}$, *where* $k_j = n_j - n_i$, $h_j = \frac{1}{n} \sum_{i \in k_j} h_i$, *and* $P = \sum \frac{k_j}{n}$.

Observe that because each individual is equally weighted in the construction of the absolute Lorenz index, then it must also be true that in the sequence of social welfare functions which are equivalent to the $AL(P)$ each individual is also equally weighted. Therefore, if the social welfare function is to be consistent with the Pigou-Dalton transfer principle, the reason for which transfers from healthier to less healthy individuals improves social welfare must be because individual utility functions are concave in health. Notice also that the equal weighting property means that a policy maker cannot, on the one hand, use the absolute Lorenz index to rank alternative distributions of population health and, on the other hand, argue that the social welfare function should put more weight on the interests of less healthy individuals.

If we are trying to compare the health of different populations, and want to gain insight into how the utilitarian social welfare functions representing each of these populations may be related to each other, it turns out to be useful to impose a non-intersection property on the $AL(P)$.

Remark 3 *(Also Shorrocks, 1983) Consider a series of absolute Lorenz indices* $AL_1(P)$, $AL_2(P), ...AL_n(P)$ *which measure inequality in the distribution of health in* n *different populations. If for all* $i, j, 0 < P < 1, AL_i(P) \cap AL_j(P) = \varnothing$ *then there exists a general social welfare function* $\omega : [W(\overrightarrow{h_1}), ..., W(\overrightarrow{h_n})] \rightarrow \mathbb{R}$, *which maps the vector of utilitarian*

social welfare functions into the set of real numbers, and such that for each element of the vector of social welfare functions $W(\overrightarrow{h_j}) = \frac{1}{\alpha_i}\sum_{i=1}^{n_j} U(h_i), \alpha_i > \alpha_j$ *if* $i < j$.

Of course, whereas this result does provide some insight into the relationship between the rankings provided by the application of the absolute Lorenz criterion and the associated social welfare functions, the practical consequences are typically likely to be limited, as the absolute Lorenz curves will generally intersect.

Just as Shorrocks (1983) were able to connect the criterion of absolute Lorenz dominance to the associated social welfare function, an equivalent exercise is to connect the rankings of alternative distributions of the health of a given population generated by the application of the stochastic dominance criterion to the rankings generated by the social welfare function which represents the underlying social preferences of the stochastic dominance criterion. What this comparison underscores is that the stochastic dominance criterion puts higher weight on the health of less healthy individuals.

Proposition 2 *When distributions of population health are ranked using stochastic dominance criteria of order greater than or equal to 2, then in the corresponding social welfare function which represents the same ordering of the populations, individuals are not equally weighted. In particular, individuals at the lower end of the distribution are more heavily weighted than are healthier individuals.*

Proof. *By definition, the second-order health dominance index of population A,* $H_A^2(z)$, *can be expressed as:*

$$H_A^2(z) = \int_0^z F(t)dt = \int_0^z (z-t)dF(t) = \int_0^z (\frac{z}{t} - 1)tdF(t).$$

This implies that in addition to the density weighting in the measurement of health, there is also an implicit weight function $\omega = (\frac{z}{t} - 1)$. *Observe that*

$$\begin{aligned} \frac{d\omega}{dt} &= -\frac{z}{t^2} < 0 \\ \frac{d^2\omega}{dt^2} &= \frac{2z}{t^3} > 0 \end{aligned}$$

and therefore ω *is a concave function of the health level. The concavity of the weight function implies that individuals who experience a lower health level are assigned relatively higher weight. This result can be easily extended to stochastic dominance of order* s : $H_A^s(z) = \int_0^z (z-t)^{s-1} dF_s(t) = \int_0^z (\frac{z}{t} - 1)(z-t)^{s-2} dF_s(t)$. ■

It is easy to observe that when s increases, the underlying social welfare function will become more and more Rawlsian, since the weight given to the lowest health level increases. Whereas this is not a serious drawback when considering issues of economic inequality, this is of significant concern when comparing that health of different populations. Health, unlike income, cannot always easily be redistributed in the population: some individuals will typically suffer from irremediable health problems, and therefore it is crucial to ensure that appropriate weight is placed on improvements in the distribution of health amongst members of the society whose health *is* susceptible to improvement if adequate resources are made available. Moreover, given that in any population there will always be individuals in very poor health, it is almost inevitable that the health of these populations will be able to be compared only using higher order stochastic dominance, and therefore these comparisons will be driven largely by the (possibly random) differences in the proportion of these individuals, rather than being sensitive to differences in the distribution

of health in the wider population that result from the thoughtful allocation of health care resources.

2.3 Mean-Inequality Measure Versus SD When Comparing the Health of Populations

In this section we relate the rankings of different distributions of population health generated by the Mean-Inequality Measure (MIM) to those generated by applying the SD criterion. Recall that the MIM is a scalar measure of the health of a population:

$$\widetilde{H} = [\alpha \overline{G}^{\rho} + \beta(\overline{H}I)^{\rho}]^{\frac{1}{\rho}}$$

where G is the average health gap of the population, that is,

$$G(F) = \int_0^{h^*} (h^* - h) dF(h)$$

and ρ is interpreted as a compensation factor that reflects the sensitivity of the measure to changes in the distribution of health and I is a measure of inequality in the distribution of health in the population. The reference set of I consists of popular indices of inequality, such as the Gini index, Atkinson Inequality Index, Theil's inequality index, coefficient of variation, ect. For the purpose of simplicity, in what follows below we use the Gini index as the measure of inequality. Recall that $Gini = 1 - 2\int_0^1 L[F(h)]dF(h)$, where L is the relative Lorenz index $RL(P)$ defined previously.

The main role of ρ in mean inequality measure is to capture the preference of the policy maker over the tradeoff between average health of the population and the health inequality of the population, representing the attitude of the policy maker towards the risk

of getting health inequality in return of health improvement in average health. And $\rho = 1$ corresponds to risk neutral, $\rho < 1$ represents risk averse.[11]

Similarly, although α and β are also parameters, it might be supposed that as the average health level increases, more concern will be given to the inequality in the health distribution. (Bleichrodt & van Doorslaer, 2006) . Observe that when $\rho = \alpha = \beta = 1$, we have the simple form of MIM. Notice that $h^* \geq \widetilde{H} \geq 0$, and that lower values of $\widetilde{H}$ are associated with higher levels of population health.

To better understand the relationship between the MIM and SD, we rewrite the MIM as:

$$\begin{aligned}
\widetilde{H} &= [\alpha \overline{G}^{\rho} + \beta(\overline{H}g)^{\rho}]^{\frac{1}{\rho}} \\
&= [\alpha[\int_0^1 (h^* - t)dF(t)]^{\rho} + \beta\left(\int_0^1 tdF(t)\left(1 - 2\frac{\int_{x=0}^1 t(1-F)dF(t)}{\int_0^1 tdF(t)}\right)\right)^{\rho}]^{\frac{1}{\rho}} \\
&= [\alpha[\int_0^1 (\frac{h^*}{t} - 1)tdF(t)]^{\rho} + \beta\left(\int_0^1 tdF(t) - 2\int_{x=0}^1 t(1-F)dF(t)\right)^{\rho}]^{\frac{1}{\rho}} \\
&= [\alpha[\int_0^1 (\frac{h^*}{t} - 1)tdF(t)]^{\rho} + \beta\left(\int_0^1 t(2F-1)dF(t)\right)^{\rho}]^{\frac{1}{\rho}}
\end{aligned}$$

Notice that individuals are weighted differently in the two arguments. In the first argument, the weight is a increasing function in the average health gap, with greater weight put on the least health individuals. In contrast, in the second argument, which reflects inequality in the distribution of health, there is relatively more weight put on healthy individuals. In

[11] when $\rho = 1$, the MIM becomes a linear function of average health and health inequality; when $\rho < 1$, the MIM is a concave function of average health and health inequality. As an extended application, MIM could also be applied to the situation of measureing individual health over periods, where ρ is used to reflect the individual risk attitude towards the inequality of individual's health across different periods.

general, SD and MIM will rank populations differently; however, this is not always true when two populations can be ranked using second-order stochastic dominance.

Proposition 3 *Suppose that two populations can be ranked using second order stochastic dominance. Then there exists* $\rho^* \leq 1$ *such that the MIM generates the same rankings as second-order SD for any* $\rho \in [\rho^*, 1]$.

Proof. See appendix. ■

Note that the range of ρ must be restricted, otherwise there could rise some extreme situation that would undermine the appropriateness of MIM.

Below is the extreme example, where population B has lower average health and higher inequality, it should be ranked as less healthier than population A as long as the MIM satisfies the reasonable axioms. However, when ρ becomes negative, $MIM_B < MIM_A$,which indicates B is healthier than A, which is clearly absurd.

$In\,\mathrm{div}\,idual$	A	B
1	0.8	0.1
2	0.8	0.1
3	0.9	0.1
4	0.9	0.1
5	0.8	0.12
$average$	0.84	0.104
$gini$	0.02875	0.03077
gap	0.16	0.896
$ave \times gini$	0.024	0.0032
$MIM_{\rho=1}$	0.184	0.8992
$MIM_{\rho=-100}$	0.024	0.0032

where $MIM = (G^\rho + (HI)^\rho)^{1/\rho}$.

This reveals the fact that the application of efficiency and equity tradeff effect by ρ should not be limitless.

It is important to consider whether or not the value of ρ can approach negative infinity. For MIM to be a meaningful distributionally-sensitive measure of population health, it must be true that when comparing two populations with the same measured level of inequality, but different average health gaps, that it ranks as healthier the population with the smaller health gap. In other words, we require that $\frac{\partial MIM}{\partial G} > 0$. Similarly, we require that when comparing two populations with the same average health gap, that it rank as healthier the population with the lower level of inequality, i.e., $\frac{\partial MIM}{\partial I} > 0$. Whereas it is straightforward to verify that $\frac{\partial MIM}{\partial I} > 0$ for all $\rho \leq 1$, the requirement that $\frac{\partial MIM}{\partial G} > 0$ in fact bounds the value of ρ strictly away from negative infinity. To establish this claim, observe that $\frac{\partial MIM}{\partial G} = (G^\rho + ((1-G)I)^\rho)^{\frac{1}{\rho}-1}(G^{\rho-1} - I^\rho(1-G)^{\rho-1})$, and therefore $\frac{\partial MIM}{\partial G} > 0$ if and only if $G^{\rho-1} - I^\rho(1-G)^{\rho-1}) > 0$,which is true only if

$$\begin{aligned} (\rho-1)\ln G &> \rho \ln I + (\rho-1)\ln(1-G) \\ \rho(\ln G - \ln I - \ln(1-G)) &> \ln G - \ln(1-G) \end{aligned}$$

and so when $\ln \frac{G}{(1-G)I} > 0$, $\frac{\partial MIM}{\partial G} > 0$ only if $\rho > \frac{\ln \frac{G}{1-G}}{\ln \frac{G}{1-G} - \ln I} = \frac{1}{1-\frac{\ln I}{\ln \frac{G}{1-G}}} = \underline{\rho}$. It can be checked that$\underline{\rho}$ is greater than the critical ρ identified in the proof of proposition 3, and therefore in this setting MIM always ranks populations in the same way as SSD, when SSD is able to generate a ranking; however, MIM provides a ranking in circumstances when SSD cannot.

Alternatively, when $\ln \frac{G}{(1-G)I} < 0$, $\frac{\partial MIM}{\partial G} > 0$ only if $\rho < \frac{1}{1-\frac{\ln I}{\ln \frac{G}{1-G}}}$. However, in this situation, $\frac{G}{(1-G)} < I$, and so $\frac{1}{1-\frac{\ln I}{\ln \frac{G}{1-G}}} > 1$.Consequently, $\frac{\partial MIM}{\partial G} > 0$ is verified for any

$\rho < 1$. In this case, when two populations can be ranked by SSD, then MIM and SSD rank the two populations similarly for all $\rho \in [\rho^*, 1]$. However, the rankings diverge for $\rho < \rho^*$. As in the first case, MIM is able to rank the two populations in settings where SSD cannot do so.

Generally speaking, as noted by Davison and Duclos (2000), although SD does not make ethical judgments over equity and efficiency and hence is amenable to a wide range of beliefs, a significant limitation of the SD approach is that it places greater weight on the least healthy members of the population, whereas the MIM places more weight on those of low to average health. Therefore, there are extreme situations where theoretically SD will not provide satisfactory answers to population health comparision while MIM can. For example, consider population A where most people are healthy except one individual (or a very small group of individuals) is in serious illness, and population B where most people have low health levels but no one is extremely unhealthy. If SD is used to compare the health of poulation A with population B, population A will be considered as less healthy and thus entitled to receive prioprity in obtaining health resources . However, it is likely the case that it is population B that is in most in need of health resources. In such a situation, the use of SD for population health comparisons will lead to health resources being directed to already healthy people. In contrast, the MIM will give a more reasonable comparison because it generates a complete ranking over the health of different populations with explicit attention given to the tradeoff between health equity and efficiency. To the extent that there may be no effective treatment available for more severe health conditions, and that such conditions may be due principally to random variation, it is in some sense

unreasonable to use a measure of the rankings of the health of different populations which may be driven principally by elements beyond the control of policy-makers. In effect, it seems desirable to select a measure of population health inequality which is responsive to those changes in the distribution of health which result from the thoughtful application of health policy. This would therefore suggest that the MIM might be preferred to SD for the purposes of comparison of the health of populations.

2.4 Does the Choice of Population Health Measure Matter In Practice?

Whereas the analysis above draws attention to the differences between the MIM and SD as inequality-sensitive measures of population health, this does not imply that they will actually rank the health of populations differently in real-world settings. In particular, if populations differ dramatically in terms of health (for example, Canada versus Nigeria), then the choice of measure may be of little importance for ranking the health of these populations, as rankings will be consistent regardless of the choice of measure. Consequently, if a strong case is to be made for the use of the MIM, it is important to show that it will in fact generate different rankings from, say, the SD measure when comparing populations with relatively similar levels of health.

2.4.1 Description of the data

For the purpose of comparing rankings generated by the MIM with those generated by the application of the SD criterion, we consider data for Canada and the United States.

Specifically, we use data collected from the Joint Canada/United States Survey of Health (JCUSH) for 2002-2003, as well as the National Population Health Survey (NPHS) for 1996-1997 and the National Health Interview Survey (NHIS) for 1996-1997.

The 2002-2003 JCUSH[12] is a cross-sectional survey administered jointly by Statistics Canada and the Centres for Disease Control in the United States, which collected health information from 3505 Canadian and 5183 U.S. residents from November 4, 2002 to March 31, 2003. The target population of the JCUSH is Canadian and American household residents aged 18 years or older. The institutionalized population is excluded, as are people in prison and full time members of the Canadian or American Armed Forces. Data was collected from respondents on the residents' health, their use of health care and their functional limitations, by using the Computer-Assisted Telephone Interviewing (CATI) method.

From this data, a Health Utility Index (HUI) score was calculated for each respondent. The HUI is a well-established measure, which is intended to provide a reasonable measure of comparison between alternative states of health. (Torrance, 1986; Feeny *et al*, 2002) This measure attains a maximum of 1, is normalized to zero for death, but can take on negative values in the case of extremely debilitating chronic conditions. In the sample collected, the minimum HUI score observed in Canada is -0.243, and the minimum HUI score in United States is -0.257.

The overall response rates of the survey are 65.5% for Canadian sample and 50.2% for the United States sample. A survey weight is given to each person included in the final sample, that is, the sample of persons who responded to the survey questions. This

---

[12] http://www.statcan.ca/cgi-bin/imdb/p2SV.pl?Function=getSurvey&SDDS=5020&lang=en&db=IMDB&dbg=f&adm=8&dis=2

weight corresponds to the number of persons represented by the respondent for the entire population. The weights for the Canadian and the U.S. samples were obtained separately, but both used the same method and the same weight adjustments.

The 1996-1997 NPHS we use here is a cross-sectional survey of the Canadian population, and collects data concerning socio-demographic characteristics and the health status of the respondents.[13] The target population of NPHS includes household residents in all provinces, with the age ranging from 0 to over 80. A sample of telephone numbers was selected from the Statistics Canada Random Digit Dialing (RDD) System. In the survey, the response rate is 82.6%, altogether there are 81,804 respondents. Data was collected from respondents via telephone interviews, personal visits and written questionnaires that were distributed via the mail. The survey includes information of health status, use of health services, risk factors and demographic and socio-economic status. A subset of the answers collected from the questionnaire were used to construct a HUI score for each respondent. The minimum HUI score in the record is 0.061, and the highest score is 1. In addition, one question was used to determine self-reported health (SRH, ranging from 1 poor health to 5 excellent health). Note that the HUI score is a relatively objective measure of health status, whereas the SRH score is largely subjective.

The NHIS is also a cross-sectional survey and, similarly to the NPHS, collects both health information and socio-economic information of the households but from United States rather than Canadian residents.[14] The interviewed sample of 1996 consists of 63,402 persons with the non-response rate 6.2%. The age of the respondents ranges from 0 to 90.

[13] http://www.statcan.ca/english/Dli/Data/Ftp/nphs/nphs1996.htm

[14] http://www.cdc.gov/nchs/about/major/nhis/quest_data_related_1969_96a.htm#1996_NHIS

Descriptive statistics of HUI scores and self reported health scores in three surveys

	N	Mean	Median	Std. Error of Mean	Minimum	Maximum	Std. Deviation
JCUSH HUI							
CANADIAN SAMPLE	1016	.88215	.95438	.006135	-.243	1.000	.195543
U.S. SAMPLE	8610	.86770	.94375	.002291	-.257	1.000	.212537
NPHS HUI							
NFLD	1541	.93409	.95382	.002746	.083	1.000	.107785
PEI	368	.92495	.95520	.006272	.098	1.000	.120374
NS	2479	.91055	.94381	.002716	.180	1.000	.135230
NB	2041	.91877	.95049	.002635	.139	1.000	.119024
QUE	19706	.92742	.95055	.000755	.083	1.000	.105042
ONT	29865	.92712	.95762	.000691	.051	1.000	.119510
MB	2916	.92212	.95495	.002252	.103	1.000	.123754
SASK	2652	.92144	.95148	.002207	.146	1.000	.113687
AB	7314	.92617	.95510	.001384	.089	1.000	.118366
BC	10202	.91881	.95038	.001178	.180	1.000	.118975
NHIS SRH							
Northeast	12807	.7927	.8222	.00183	.20	1.00	.20702
Midwest	15070	.7910	.8182	.00165	.20	1.00	.20438
South	21448	.7697	.7997	.00151	.20	1.00	.22094
West	13379	.7883	.8175	.00182	.20	1.00	.21071
NPHS SRH							
Atlantic	5585	.76337	.77391	.002402	.200	1.000	.194882
Central	51088	.77550	.79312	.000864	.200	1.000	.195204
Prairies	13600	.75972	.78409	.001692	.200	1.000	.197373
Pacific	10528	.76962	.78302	.001893	.200	1.000	.194224

Data was collected from respondents via telephone interviews and written questionnaires that were distributed via the mail. The data includes the detail information about demographic and socio-economic status, health status, health service utilization, access to care are, etc. Respondents are asked to report on their health status, by ranking their health on a scale of 1 to 5, where 1 is poor health and 5 is excellent health. 2.7% of the respondents are recorded as being in poor health while 35.8% claim to have excellent health.

The brief health information in the three surveys are summarized in the above table.

2.4.2 Country-level Comparisons of the Health Status of the Canadian and United States Populations

Our first comparison exercise uses the data from the JCUSH survey to compare the health of the Canadian and United States populations. We construct rankings of these two populations using a variety of different measures: the mean of the HUI scores; first and second order stochastic dominance, the Gini coefficient, and the MIM. The results of this exercise are reported in table 1.

Table 1 health comparison between Canada and US (HUI)

	mean	ranking	SD1	SD2	gini	MIMranking
Canada	0.8821	1	?	1	0.093	1
US	0.8676	2	?	2	0.106	2

Notes: ? indicates ambiguous ranking, 1 implies ranking higher than 2.

The mean of the HUI scores for Canada is 0.8821, whereas for the United States it is only 0.8676. Using the Kolmogorov test, we reject the null hypothesis that the means are the same at the 95% confidence level, and conclude that the population health ranking should favour Canada. We subsequently compare the health of these two populations using the first and second degree stochastic dominance criterion. To this end, we consider the distribution of the health of individuals falling below successively higher HUI scores, and calculate the mean of the HUI scores for these successive sub-populations. Taking into consideration the sampling error, we construct t-values enabling us to test the null hypothesis that the mean of the HUI scores is the same in the two populations when restricting attention only to those individuals belong to the particular sub-population under consider-

ation. Note that $t = \frac{H_A^s - H_B^s}{\sqrt{var(H_A^s)+var(H_B^s)}}$.[15] If all of the t values are negative (positive), then A stochastically dominates B (B stochastically dominates A) at order s.[16] In the JCUSH data set, the t-values are negative (with Canada having a lower cumulative distribution than the United States) up until a health level of 0.973, at which point the t-values become positive. This means that the two populations cannot be ranked according to the first-order stochastic dominance criterion.

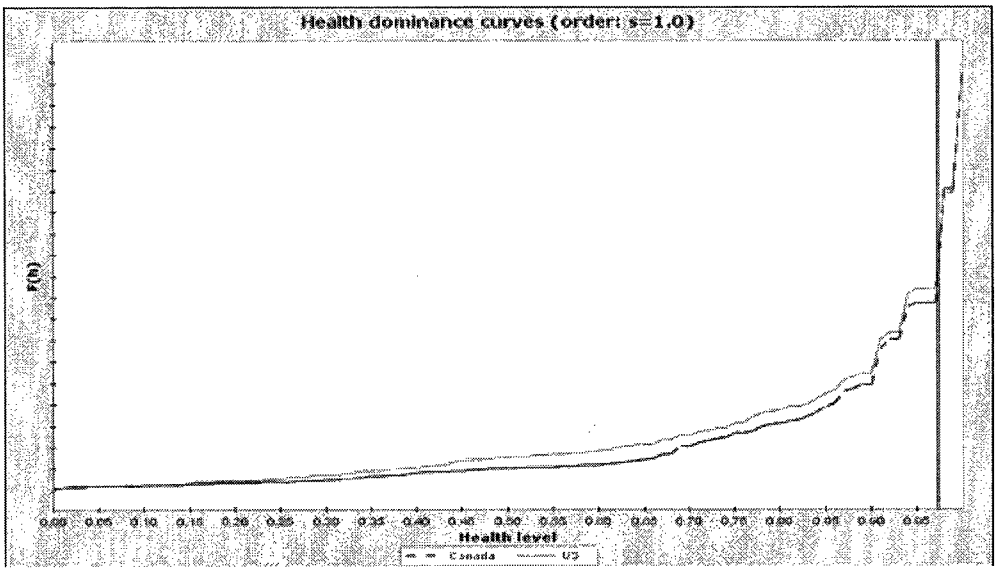

Fig. 2.7. Stochastic dominance curves at order 1

Next, we compare the data using the second-order stochastic dominance criterion. This involves calculating the cumulative health gap for successive sub-populations. Once again, we calculate the appropriate t-values, and as these are consistently negative, enabling

[15] The estimation process of H^s and $var(H^s)$ could be found in Davidson&Duclos (1998), where they link the FGT to these estimators.

[16] The null hypothesis also could take the equality form, $H_0 : H_A^s = H_B^s$, in case where the null is rejected and the signs are the same on all of the t statistics, then dominance of order s is declared. Please refer to David Stifel's course notes. http://ww2.lafayette.edu/~stifeld/. In this situation, when the signs of t changes, a crossing point of two distributions is specified, but the significance has to be checked in order to claim difference between the two distributions.

us to conclude that Canada is healthier than the United States according to the criterion of second-order stochastic dominance.

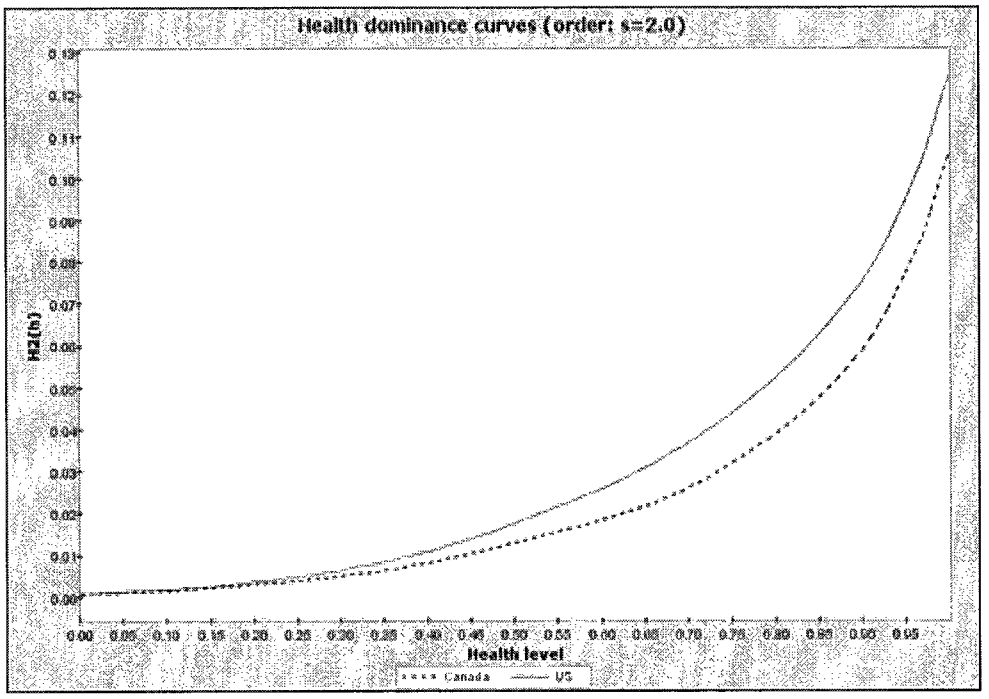

Fig. 2.8. Stochastic dominance curves at order 2

Whereas the Gini coefficient does not provide an aggregate measure of population health, it does provide insight into differences in the distribution of health across populations. If we calculate the Gini index for the distribution of the HUI scores for Canada and the United States, we find that Canada's Gini score is 0.093, whereas the United States has a score of 0.106, indicating that health in Canada is more equally distributed than in the United States.

Finally, we use the MIM to compare the two populations. For simplicity, we let $\alpha = \beta = 1$ in the MIM, and consider a range of different values for $\rho : 0 < \rho < 1$, and $-\infty < \rho < 0$. Simulation results show that, regardless of the value of ρ, Canada

has a lower MIM score than the United States, which means that Canada's population is considered to be healthier.

Here Canada is consistently ranked as healthier than US; given the results of Proposition 3, this is no surprise.

2.4.3 Health comparisons between Canada and the United States (regional level)

The above health comparison is based on the population as a whole. However, the health policy maker might be interested in putting the comparison in a context where relatively different regions in a country or in a commonwealth in which the member countries share similar socioeconomic characteristics, in order to explore the spatial variation of health determinants. Because the JCUSH data set doesn't provide regional breakdowns, we turn to the NPHS for 1996-1997 (Canada) and the NHIS for 1996 (US), mainly due to their availability and comparability. Though there is no HUI measure in 1996 NHIS, both data sets contain self-reported health as a measure of individual health and the coding is the same.

We can compare the health of these regions, using the different tools discussed above. The region wise rankings are recorded in Table 2, the health distributions of the regions are ordered according to the transitivity rule. Since the interpretation of stochastic dominance results for orders greater than 3 is difficult, we limit the use of SD to orders less than or equal to three, and so record the ranking as ambiguous if regions cannot be classified using SD of order 3 or below. From the table, we note that the regions are essentially divided into 3 groups. The top regions are Canada Central and the Pacific Region; the performance

of these regions is quite evident, and not surprising. The middle-ranked group contains the Canada Atlantic region and US Northeast and Midwest regions. These regions have quite similar distributions of health outcomes, and therefore it is not surprising that their rankings relative to each other are ambiguous. The lowest ranked group consists of the Canada Prairies region and US South and West regions; however, the US South region is dominated by all other regions.

Table 2 Regional health comparison (1996 SRH) Stochastic dominance

Region	Northeast	Midwest	South	West	Atlantic	Central	Prairies	Pacific	ranking
Northeast		≥?	≥(1)	≥(2)			≥(2)		4
Midwest			≥(1)	≥(3)			≥(2)		5
South									8
West			≥(1)						7
Atlantic	≥?	≥?	≥(2)	≥(3)			≥(2)		3
Central	≥(1)	≥(1)	≥(1)	≥(1)	≥(1)		≥(1)	≥(2)	1
Prairies			≥(3)	≥?					6
Pacific	≥(3)	≥(3)	≥(1)	≥(2)	≥(2)		≥(1)		2

notes: "$\geq$" indicates stochastic dominance running from row region to column region, and the number afterwards specifies the dominance order, "?" implies a order higher than 3.

Table 3 presents ranks these same regions using the MIM criterion rather than SD. We observe that when $-1.3 < \rho < -1.1$, which means the elasticity of substitution ranges 0.434-0.476, then the rankings of the regions are relatively consistent with SD. However, outside this interval, the rankings generated by the MIM and the SD criterion differ. In particular, when $0 < \rho < 1$, the MIM calculation weights more heavily the argument that is less equally distributed. In the data sets considered here, the range of mean health is

much larger than the range of the Gini coefficient. Consequently, when ρ is positive, the Northeast region and Midwest region are ranked above all other regions, since they have the highest mean health levels. In contrast, when ρ is negative, the MIM is less sensitive to differences in the variance of the distribution of the mean health level versus the Gini coefficient.

Table 3 regional health comparison (1996 SRH) MIM

	Region	Average	Ranking	Gini	Ranking	ρ=-1.3	ranking	ρ=-1.2	ranking	ρ=-1.1	ranking
US	Northeast	0.792679	1	0.139694	6	0.083536	5	0.080271	5	0.076517	4
	Midwest	0.791043	2	0.138408	5	0.08308	4	0.079853	3	0.076139	2
	South	0.769737	5	0.155284	8	0.090954	8	0.087433	8	0.083377	8
	West	0.788349	3	0.143425	7	0.085294	7	0.08196	7	0.078127	7
Canada	Atlantic	0.763365	8	0.136982	3	0.083207	3	0.08018	4	0.076651	5
	Central	0.775501	4	0.135108	1	0.082181	1	0.079123	1	0.075573	1
	Prairies	0.768719	7	0.138374	4	0.083761	6	0.080662	6	0.077062	6
	Pacific	0.769622	6	0.136171	2	0.082765	2	0.079716	2	0.076172	3

2.4.4 Health comparison across provinces of Canada

The third exercise consists in using both SD and the MIM to rank the health of the populations in each of the ten Canadian provinces. These provinces are characterized by roughly comparable health systems and socioeconomic characteristics. Not surprisingly, therefore, there is not a lot of variation across provinces in the average health status. The results of the comparisons are presented in Tables 4 and 5. Observe that the provinces can be ranked us-

ing SD of order no greater than 2. Additionally, since higher average health in a province is usually accompanied by lower inequality, the MIM ranks the provinces in a way similar to SD for a large range of values of ρ.

Table 4 health comparison across provinces of Canada (1996-1997 HUI) stochastic dominance

Region	NFLD	PEI	NS	NB	QUE	ONT	MB	SASK	AB	BC	ranking
NFLD		≥(1)	≥(1)	≥(1)	≥(1)	≥(1)	≥(1)	≥(1)	≥(1)	≥(1)	1
PEI			≥(1)	≥(2)	≤(2)	≤(2)	≥(2)	≥(1)	≤(2)	≥(1)	5
NS				≤(2)	≤(2)	≤(1)	≤(1)	≤(2)	≤(1)	≤(2)	10
NB					≤(2)	≤(1)	≤(1)	≤(2)	≤(1)	≤(2)	9
QUE						≥(2)	≥(2)	≥(2)	≥(2)	≥(1)	2
ONT							≥(2)	≥(2)	≥(2)	≥(1)	3
MB								≤(2)	≤(2)	≥(2)	7
SASK									≤(2)	≥(2)	6
AB										≥(1)	4
BC											8

Notes: "$\geq$" indicates stochastic dominance running from row region to column region, and the number afterwards specifies the dominance order.

The results reported in table 5 show that there is considerable variation in the distribution of health in the Atlantic region. Newfoundland is ranked as the healthiest province in Canada, whereas Nova Scotia and New Brunswick are the least healthy. As might be expected, Quebec and Ontario rank high in the list, followed by Alberta, PEI and the other two prairie provinces. It is somewhat surprising that BC is ranked in eighth place, whereas it was ranked either second or third in table 3. This may be due to the fact that table 5 uses HUI, whereas table 3 relies on self-reported health status.

Table 5 health comparison across provinces of Canada (1996-1997 HUI) MIM

Region	Average	ranking	gini	ranking	ρ=0.9	ranking	ρ=0.2	ranking	ρ=-0.5	ranking	ρ=-1.3	ranking
NFLD	0.93409	1	0.04762	1	0.118995	1	1.739257	1	0.013406	1	0.030987	1
PEI	0.92496	5	0.05435	5	0.135072	5	1.973233	5	0.015201	5	0.035117	5
NS	0.91065	10	0.06496	10	0.160054	10	2.336268	10	0.017983	10	0.041505	10
NB	0.91877	9	0.05677	9	0.143722	9	2.093158	9	0.016075	9	0.037009	9
QUE	0.92742	2	0.04878	2	0.126913	2	1.844035	2	0.014128	2	0.032442	2
ONT	0.92712	3	0.05334	4	0.131887	3	1.92841	3	0.014869	3	0.034384	3
MB	0.92212	6	0.05651	8	0.140119	7	2.046814	7	0.015767	7	0.036421	7
SASK	0.92144	7	0.05471	6	0.138956	6	2.023625	6	0.01554	6	0.035775	6
AB	0.92617	4	0.05317	3	0.13266	4	1.937416	4	0.014921	4	0.034458	4
BC	0.91981	8	0.05596	7	0.141857	8	2.065923	8	0.015865	8	0.036524	8

2.5 Conclusion

This paper investigates the relationship between different approaches to ranking the health of populations, and specifically investigates the links between Lorenz dominance, stochastic dominance, and the mean-inequality measure. The theoretical analysis shows that, in general, these measures will rank the health of populations differently. In particular, rankings generated using stochastic dominance will generally be highly sensitive to the number of individuals reporting very low levels of health, whereas the rankings generated by the mean-inequality measure are more sensitive to differences in the variance of the distributions. To the extent that low levels of health may be largely - if not entirely - due to factors

that lie beyond the control of policy makers (for example, genetic factors), ranking the health of populations by applying the stochastic dominance criterion would seem to be somewhat less compelling than is the case in the analysis of poverty. In contrast, to the extent that the thoughtful application of health policy can reduce variance in the distribution of health, as well as average health levels, the MIM seems to be a serious candidate for consideration when ranking the health of different populations.

Whereas the theoretical analysis of these measures suggests that they may generate different rankings, it is clearly of interest to see whether or not these two approaches generate significantly different rankings in actual practical applications. To this end, we use both stochastic dominance and the mean-inequality measure to compare the health of different populations using three different data sets. Whereas both SD and MIM rank the Canadian population higher than the USA population, they differ when used to rank the health of different regions.

2.A Proof of Remark 1

Proof. As illustrated in figure 2A.1, consider two populations A and B, with health distributions $F_A(h)$ and $F_B(h)$, where these distributions do not intersect. For simplicity, assume that B lies above A.Notice that for any $h \in (0, \min[F_A^{-1}(1), F_B^{-1}(1)])$, $F_B(h) > F_A(h)$. Figure 2A.1 can be translated into an equivalent "Lorenz setting", through the intermediary of inverse cumulative distribution curves, where the y axis and x axis are interchanged. This generates the symmetric graphs for figure 2A.1, (see figure 2A.2).

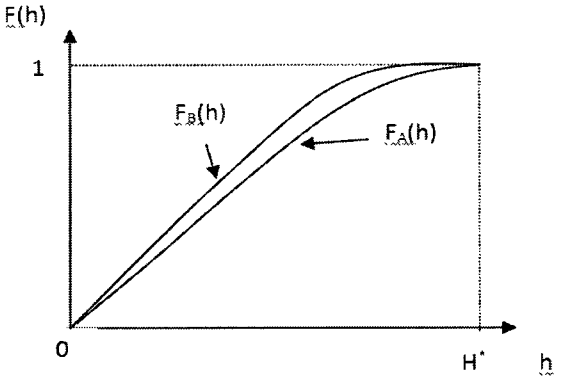

Figure 2A.1

Inverse cumulative distribution (ICD) curve reflects the same relationship between health and its distribution, and accordingly there is no intersection between the curves. The difference is that the ICD of B lies below that of A.

Recall that when integrating $F^{-1}(P)$ (the ICD function) over $F(h)$, we obtain the absolute Lorenz index $AL(P) = \int_0^{F^{-1}(P)} h dF(h)$,which is the size of the area under $F^{-1}(P)$ from 0 till P^*(see Figure 2A.3).

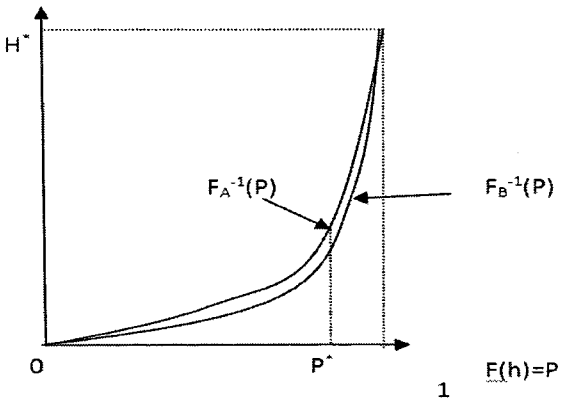

Figure 2A.2

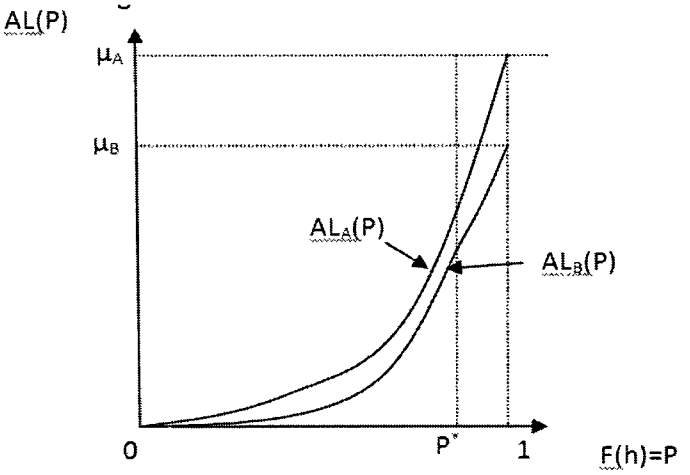

Figure 2A.3

If $F_A^{-1}(P) \geq F_B^{-1}(P)$ for any P, and $F_A^{-1}(P) > F_B^{-1}(P)$ for at least one P, then $AL_A(P) \geq AL_B(P)$ for any P and $AL_A(P) > AL_B(P)$ for some P, which means that A absolute Lorenz dominates B. However, observe that $F_A^{-1}(P) \geq F_B^{-1}(P)$ implies that $F_A^{-1}(P) = F_B^{-1}(P) + a$, $a \geq 0$.

Since the cumulative distribution function is non-decreasing, $F_A[F_A^{-1}(P)] == P == F_B[F_B^{-1}(P)] \leq F_B[F_B^{-1}(P) + a]$. This is equivalent to the condition that $F_A(h) \leq F_B(h)$

for any $h = F_A^{-1}(P) = F_B^{-1}(P) + a$, with strict inequality holds for some h, but this is simply the non-intersection condition of the health distribution curves.

Observe that the non-intersection condition implies the dominance of first moments, due to the non-decreasing property of the Lorenz index. In the example above, we have $\mu_A > \mu_B$. The non-intersection condition of the health distribution curves is sufficient to ensure that there is Absolute Lorenz Dominance; a comparison of the first moments establishes which distribution dominates the other one. ■

2.B Proof of proposition 1

Proof. According to Atkinson (1970), second order stochastic dominance is equivalent to Absolute Lorenz dominance. If two Absolute Lorenz curves intersect, the corresponding stochastic dominance curves must also intersect. (Chiu, 2007, Proposition 1) proves that when Lorenz curve of A crosses Lorenz curve B first n time from above is equivalent to the corresponding stochastic dominance curve of B crosses that of A n time from above. As a special case, if there is a single intersection of Lorenz curves, there is also intersection in the related stochastic dominance curves, but the order of crossing is opposite.

Necessity: Given that the second order stochastic dominance curve of the health distribution of B crosses that of A from above, if third order dominance holds, the dominance curves of B at order 3 has to be above that of A for all h, it indicates that , from the definition, we have at the highest level of health H^*,

$$\frac{1}{n}\sum_{i=1}^{n}(H^* - h_i)_B^2 \geq \frac{1}{n}\sum_{i=1}^{n}(H^* - h_i)_A^2 \tag{2.2}$$

Because

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(H^* - h_i)^2 &= \frac{1}{n}\sum_{i=1}^{n}(H^* - \bar{h} + \bar{h} - h_i)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}(H^* - \bar{h})^2 + \frac{1}{n}\sum_{i=1}^{n}(\bar{h} - h_i)^2 \\ &= \bar{g}^2 + \sigma^2\end{aligned}$$

where $\bar{g} = H^* - \bar{h}$, therefore, (2.2) gives us that $\bar{g}_B^2 + \sigma_B^2 \geq \bar{g}_A^2 + \sigma_A^2$,After rearrangement, we have $\sigma_B^2 - \sigma_A^2 \geq \bar{g}_A^2 - \bar{g}_B^2$

Alternatively, $\sigma_B^2 - \sigma_A^2 = (\sigma_B + \sigma_A)(\sigma_B - \sigma_A)$, $\bar{g}_A^2 - \bar{g}_B^2 = (\bar{g}_A + \bar{g}_B)(\bar{g}_A - \bar{g}_B)$ which directly leads to $\frac{\sigma_B - \sigma_A}{\bar{h}_B - \bar{h}_A} \geq \frac{\bar{g}_B + \bar{g}_A}{\sigma_B + \sigma_A}$.

Sufficiency: Given that the second order stochastic dominance curve of the health distribution of B crosses that of A from above, if $\sigma_B^2 - \sigma_A^2 \geq \bar{g}_A^2 - \bar{g}_B^2$, then it implies that at the highest health level H^*, we have 2.2., which means that is no crossing point at the third order stochastic dominance curves, otherwise there is contradiction. Consequently, B dominates A at order 3. ■

2.C Proof of Proposition 3

Proof. Consider two populations, A and B. By assumption, these populations can be ranked using second order stochastic dominance (SSD); for convenience, assume that $B >_{sd}^2 A$. Then, from the definition of stochastic dominance, it must be true that $\bar{g}_A \geq \bar{g}_B$, which

means population A has higher average health gap. Observe that the MIM rankings depend on both the average health gap ($\bar{g}$), and the degree of inequality(I). Bacically, there are two cases: 1) $\bar{g}_A \geq \bar{g}_B$ and $I_A \leq I_B$; 2) $\bar{g}_A \geq \bar{g}_B$ and $I_A \geq I_B$.

The structure of the proof is as follows: first, we show that if $\rho = 1$ then $B >^2_{sd} A \Rightarrow MIM_A \geq MIM_B$. Next, for the case where $I_A \leq I_B$, we show that there exists a ρ^* such that $MIM_A = MIM_B$ for $\rho = \rho^*$. Then, we show that for any $\rho > \rho^*, MIM_A \geq MIM_B$ whereas for $\rho < \rho^*, MIM_A \leq MIM_B$. Subsequently, we we establish similar results for the case where $I_A \geq I_B$. This last step completes the proof. ■

Lemma 4 *Assume $B >^2_{sd} A$. Then $MIM_A \geq MIM_B$ when $\rho = 1$.*

Proof. For simplicity, we use the Gini coefficient as the measure of health inequality in the MIM. Recall that if $\bar{g}_A \geq \bar{g}_B$, then $\bar{h}_A \leq \bar{h}_B$. By definition, $MIM_A = [(\bar{g}_A)^\rho + (H_A I_A)^\rho]^{\frac{1}{\rho}}$. Now, observe that when $\rho = 1$, we can express MIM_B as:

$$MIM_B = (\bar{g}_A - \Delta g) + (\bar{h}_A + \Delta g)(I_A(\overset{-}{+})\Delta I).$$

Observe that if $I_A \geq I_B$ then the last expression becomes $(I_A - \Delta I)$ and therefore $MIM_A \geq MIM_B$, since $MIM_B = \bar{g}_A + \bar{h}_A I_A - \Delta g(1 - I_A + \Delta I) - \bar{h}_A \Delta I \leq MIM_A$. Suppose instead that inequality is larger in B. The last expression then becomes $(I_A + \Delta I)$. Given that, by assumption, $B >^2_{sd} A$, then if distribution B generalized Lorenz dominates distribution A it must be the case that the health levels of all persons in population B are at least as high as the health levels of identically-ranked individuals in population A. Consequently, if population B exhibits both higher average health and higher inequality than population A, then the worst possible case is where all of the increase in average health -

and all of the increase in inequality - is attributable to increased health of the healthiest individual. In other words, if h_n denotes the health of the healthiest individual in population A, then the health of that individual in population B can be expressed as $h_n + n\Delta g$, where Δg measures the increase in the average level of health. In this case, we can express the measure of health inequality in populations A and B as:

$$I_A = 1 + \frac{1}{n} - \frac{2(h_n + 2h_{n-1} + 3h_{n-2} + ... + nh_1)}{n\sum h_i},$$

$$I_B = 1 + \frac{1}{n} - \frac{2(h_n + n\Delta g + 2h_{n-1} + 3h_{n-2} + ... + nh_1)}{n(\sum h_i + n\Delta g)}.$$

Observe that

$$\begin{aligned}
\Delta I &= I_B - I_A \\
&= \frac{2(h_n + 2h_{n-1} + 3h_{n-2} + ... + nh_1)}{n\sum h_i} - \frac{2(h_n + n\Delta g + 2h_{n-1} + 3h_{n-2} + ... + nh_1)}{n(\sum h_i + n\Delta g)} \\
&= \frac{2\Delta g(h_{n-1} + 2h_{n-2} + ... + (n-1)h_1)}{(\sum h_i)(\sum h_i + n\Delta g)}
\end{aligned}$$

Therefore,

$$\begin{aligned}
MIM_A - MIM_B &= \Delta g(1 - I_A) - \Delta I(\bar{h}_A + \Delta g) \\
&= \Delta g(1 - I_A - 2(\frac{(h_{n-1} + 2h_{n-2} + ... + (n-1)h_1)}{(\sum h_i)(\sum h_i + n\Delta g)})(\frac{\sum h_i}{n} + \Delta g)) \\
&= \Delta g(1 - I_A - \frac{2}{n}(\frac{(h_{n-1} + 2h_{n-2} + ... + (n-1)h_1)}{(\sum h_i)}) \\
&= \Delta g(1 - 1 - \frac{1}{n} + \frac{2(h_n + 2h_{n-1} + 3h_{n-2} + ... + nh_1)}{n\sum h_i} - \\
&\quad \frac{2}{n}(\frac{(h_{n-1} + 2h_{n-2} + ... + (n-1)h_1)}{(\sum h_i)}) \\
&= \Delta g(\frac{2(\sum h_i)}{n\sum h_i} - \frac{1}{n}) = \Delta g\frac{1}{n} = \frac{\Delta g}{n} \geq 0.
\end{aligned}$$

Since this inequality is true even when all of the gain in average health is attributable solely to gains to the healthiest individual, it follows that this relationship is true also when this gain is attributable to less-healthy individuals, e.g., to individuals in the top-half of the health distribution. Consequently, if $B >^2_{sd} A$ then $MIM_A - MIM_B \geq 0$ when $\rho = 1$.

We now show that if $\bar{g}_A \geq \bar{g}_B$ and $I_A \leq I_B$, there exists $\rho^* < 1$ such that $MIM_A \geq MIM_B$ if and only if $\rho \in [\rho^*, 1]$. In other words, the rankings generated by MIM and by SSD coincide only if $\rho \in [\rho^*, 1]$. ■

Lemma 5 *Assume $B >^2_{sd} A$. There exists $\rho^* < 1$ such that $MIM_A = MIM_B$ for $\rho = \rho^*$.*

Proof. When $\rho < 1$, then $MIM_A = [(\bar{g}_A)^\rho + (H_A I_A)^\rho]^{\frac{1}{\rho}}$. Without loss of generality, assume that $\bar{g}_A > \bar{g}_B$ and $I_A < I_B$ Then it is possible to express MIM_B as:

$$MIM_B = \{(\bar{g}_A - \Delta g)^\rho + [(\bar{h}_A + \Delta g)(I_A + \Delta I)]^\rho\}^{\frac{1}{\rho}}$$

We need to show that for any $(\Delta g, \Delta I)$ such that $\bar{g}_A > \bar{g}_B$ and $I_A < I_B$, there exists a $\rho^* < 1$ such that $MIM_A = MIM_B$ for $\rho = \rho^*$. We proceed by considering two extreme cases: in the first case, ΔI is maximal, and in the second case, ΔI is minimal, i.e., there is only a slight difference in inequality across the two populations. Given that there exists a $\rho^* < 1$ such that $MIM_A = MIM_B$ for $\rho = \rho^*$ for each of these extreme cases, it is then immediate that this is also true for intermediate values of ΔI.

For a given increase in average health, Δg, the maximal increase in inequality occurs when all of that increase in average health level is attributable to a gain in health on the part of the healthiest individual. In this case, the last term in the expression for MIM_B becomes $(\bar{h}_A + \Delta g)(I_A + \Delta I) = \bar{h}_A I_A + \Delta g(1 - \frac{1}{n})$. By first order approximation, we

then have

$$MIM_B = \{\bar{g}_A^{\rho} - \rho\bar{g}_A^{\rho-1}\Delta g + (\bar{h}_A I_A)^{\rho} + \rho(\bar{h}_A I_A)^{\rho-1}\Delta g(1 - \frac{1}{n})\}^{\frac{1}{\rho}}$$

whereas

$$MIM_A = [(\bar{g}_A)^{\rho} + (H_A I_A)^{\rho}]^{\frac{1}{\rho}}.$$

Equating MIM_A and MIM_B and solving for ρ^*, we obtain that

$$\rho^* = 1 - \frac{\ln(1-\frac{1}{n})}{\ln \bar{h}_A I_A - \ln \bar{g}_A}, \text{where} \frac{\ln(1-\frac{1}{n})}{\ln \bar{h}_A I_A - \ln \bar{g}_A} > 0.$$

Next, suppose ρ' is a convex combination of ρ^* and 1,i.e. $\rho' = \lambda\rho^* + (1-\lambda)1 = 1 - \lambda\frac{\ln(1-\frac{1}{n})}{\ln \bar{h}_A I_A - \ln \bar{g}_A}$. Recall that $\bar{g}_A = \bar{h}_A I_A (1-\frac{1}{n})^{\frac{1}{\rho-1}}$ at ρ^*, and therefore at ρ',

$$\begin{aligned} MIM_B &= \left[\bar{g}_A^{\rho'} + (\bar{h}_A I_A)^{\rho'} + \rho'(\bar{h}_A I_A)^{\rho'-1}\Delta g(1-\frac{1}{n}) - \rho'\bar{g}_A^{\rho'-1}\Delta g\right]^{1/\rho'} \\ &= \left[\bar{g}_A^{\rho'} + (\bar{h}_A I_A)^{\rho'} + \rho'\Delta g\left[(\bar{h}_A I_A)^{\rho'-1}(1-\frac{1}{n}) - \bar{g}_A^{\rho'-1}\right]\right]^{1/\rho'} \end{aligned}$$

Given that $\rho' = 1-\lambda\frac{\ln(1-\frac{1}{n})}{\ln \bar{h}_A I_A - \ln \bar{g}_A}$, we observe that $(\bar{h}_A I_A)^{\rho'-1}(1-\frac{1}{n}) = (\bar{h}_A I_A)^{-\lambda\frac{\ln(1-\frac{1}{n})}{\ln \bar{h}_A I_A - \ln \bar{g}_A}}(1-\frac{1}{n})$ and $\bar{g}_A^{\rho'-1} = (\bar{h}_A I_A)^{-\lambda\frac{\ln(1-\frac{1}{n})}{\ln \bar{h}_A I_A - \ln \bar{g}_A}}(1-\frac{1}{n})^{\lambda}$. Now, since $0 < \lambda < 1$, it is the case that $(\bar{h}_A I_A)^{-\lambda\frac{\ln(1-\frac{1}{n})}{\ln \bar{h}_A I_A - \ln \bar{g}_A}}(1-\frac{1}{n})^{\lambda} > (\bar{h}_A I_A)^{-\lambda\frac{\ln(1-\frac{1}{n})}{\ln \bar{h}_A I_A - \ln \bar{g}_A}}(1-\frac{1}{n})$ and therefore MIM_B is less than MIM_A for any ρ'.

Next consider the second extreme case where the increase in inequality which accompanies the increase in average health is minimized. Hence, $MIM_B = \{(\bar{g}_A - \Delta g)^{\rho} + [(\bar{h}_A + \Delta g)(I_A + \Delta I)]^{\rho}\}^{\frac{1}{\rho}} = \{(\bar{g}_A - \Delta g)^{\rho} + [(\bar{h}_A I_A + \Delta g I_A + \bar{h}_A \Delta I + \Delta g \Delta I)]^{\rho}\}^{\frac{1}{\rho}}$,

In this case, by first order approximation,

$$\begin{aligned} MIM_B &= \{\bar{g}_A^{\rho} - \rho\bar{g}_A^{\rho-1}\Delta g + (\bar{h}_A I_A)^{\rho} + \rho(\bar{h}_A I_A)^{\rho-1}(\Delta g I_A + \bar{h}_A \Delta I + \Delta g \Delta I)\}^{\frac{1}{\rho}} \\ &= \{\bar{g}_A^{\rho} - \rho\bar{g}_A^{\rho-1}\Delta g + (\bar{h}_A I_A)^{\rho} + \rho(\bar{h}_A I_A)^{\rho-1}\Delta g(I_A + \Delta I + \bar{h}_A \frac{\Delta I}{\Delta g})\}^{\frac{1}{\rho}} \end{aligned}$$

We set $I_A + \Delta I + \bar{h}_A \frac{\Delta I}{\Delta g} = C$, and solve for ρ^* by equating MIM_A and MIM_B. This yields

$$\rho^* = 1 - \frac{\ln C}{\ln \bar{h}_A I_A - \ln \bar{g}_A}$$

as a special case, when $\Delta I = 0$,

$$\rho^* = \frac{\ln \bar{g}_A - \ln \bar{h}_A}{\ln \bar{g}_A - \ln \bar{h}_A - \ln I}$$

Once again, it is then straightforward to show that for any $\rho' = \lambda\rho^* + (1-\lambda)1$, MIM generates the same ranking as SSD.

Finally, following a similar procedure, we now show that if $\bar{g}_A \geq \bar{g}_B$ and $I_A \geq I_B$, there exists $\rho^* < 1$ such that $MIM_A \geq MIM_B$ if and only if $\rho \in [\rho^*, 1]$. In other words, the rankings generated by MIM and by SSD coincide only if $\rho \in [\rho^*, 1]$.

In this case, MIM_B can be expressed in terms of MIM_A

$$MIM_B = \{(\bar{g}_A - \Delta g)^\rho + [(\bar{h}_A + \Delta g)(I_A - \Delta I)]^\rho\}^{\frac{1}{\rho}}$$

We again proceed by considering two extreme cases: in the first case, ΔI is maximal, and in the second case, ΔI is minimal. ΔI is maximal in the case when all of the increase in the average health gap accrues to the least healthy individual. In this case,

$$\Delta I = I_A - I_B = \frac{2\Delta g(n\sum h_i - h_n + 2h_{n-1} + 3h_{n-2} + ... + nh_1)}{(\sum h_i)(\sum h_i + n\Delta g)}.$$

Observe that the final term in the expression for MIM_B becomes $(\bar{h}_A + \Delta g)(I_A + \Delta I) = \bar{h}_A I_A + \Delta g(\frac{1}{n} - 1)$. By first order approximation, we then have that

$$MIM_B = \{\bar{g}_A^\rho - \rho\bar{g}_A^{\rho-1}\Delta g + (\bar{h}_A I_A)^\rho - \rho(\bar{h}_A I_A)^{\rho-1}\Delta g(1 - \frac{1}{n})\}^{\frac{1}{\rho}}.$$

Equating MIM_A and MIM_B and solving for ρ^*, we obtain that

$$\rho^* = 1 - \frac{\ln(1-\frac{1}{n})}{\ln \bar{h}_A I_A + \ln \bar{g}_A}, \text{where} \frac{\ln(1-\frac{1}{n})}{\ln \bar{h}_A I_A + \ln \bar{g}_A} > 0.$$

Suppose ρ' is a convex combination of ρ^* and 1, then it is easy to show that MIM_B is less than MIM_A for any ρ'.

Alterntively, consider the situation where the decrease in inequality which accompanies the increase in average health is minimized. Hence, $MIM_B = \{(\bar{g}_A - \Delta g)^\rho + [(\bar{h}_A + \Delta g)(I_A - \Delta I)]^\rho\}^{\frac{1}{\rho}} = \{(\bar{g}_A - \Delta g)^\rho + [(\bar{h}_A I_A + \Delta g I_A - \bar{h}_A \Delta I - \Delta g \Delta I)]^\rho\}^{\frac{1}{\rho}}$,

In this case, by first order approximation,

$$\begin{aligned} MIM_B &= \{\bar{g}_A^\rho - \rho \bar{g}_A^{\rho-1} \Delta g + (\bar{h}_A I_A)^\rho + \rho(\bar{h}_A I_A)^{\rho-1}(\Delta g I_A - \bar{h}_A \Delta I - \Delta g \Delta I)\}^{\frac{1}{\rho}} \\ &= \{\bar{g}_A^\rho - \rho \bar{g}_A^{\rho-1} \Delta g + (\bar{h}_A I_A)^\rho + \rho(\bar{h}_A I_A)^{\rho-1} \Delta g (I_A - \Delta I - \bar{h}_A \frac{\Delta I}{\Delta g})\}^{\frac{1}{\rho}} \end{aligned}$$

We set $I_A - \Delta I - \bar{h}_A \frac{\Delta I}{\Delta g} = C'$, and solve for ρ^* by equating MIM_A and MIM_B. This yields

$$\rho^* = 1 - \frac{\ln C'}{\ln \bar{h}_A I_A - \ln \bar{g}_A}$$

as a special case, when $\Delta I = 0$,

$$\rho^* = \frac{\ln \bar{g}_A - \ln \bar{h}_A}{\ln \bar{g}_A - \ln \bar{h}_A - \ln I}.$$

Once again, we observe that for any convex combination of ρ^* and 1, $\rho' = \lambda \rho^* + (1-\lambda)1$, MIM generates the same ranking as SSD. ■

Chapter 3
Measurement of Population Health Over Time

3.1 Introduction

In this chapter we extend our study of distribution sensitive measures of population health to take account of the specific conceptual challenges posed by the fact that individual health outcomes change over time. There are a number of issues which make it particularly important to take account of changes over time in the distribution of health. The first, and most obvious, is the issue of preventive health care, and how it affects future health outcomes. If a static comparison of the distributions of health in two countries is undertaken, and one country is investing resources in preventive care, whereas the other isn't, then at the time at which the comparison is undertaken it may be the case that health outcomes in the two countries are roughly comparable, but the country undertaking the preventive health measures can be expected to have better outcomes over time. Clearly, only by developing a dynamic measure of population health it is possible to better capture the way in which the health of these populations differ. The second issue is that health, arguably even more so than other variables such as income, is subject to random fluctuations - for example, associated with the outbreak of specific epidemics - and therefore it is clear that we are much more interested in the distribution of health levels for individuals over time, rather than by the distribution of health levels at any specific moment in time. Finally, and most

importantly of all, because health - unlike cash - cannot be directly redistributed from one individual to another, then to the extent that health policy affects the distribution of health it is clearly necessary to develop measures of population health that take appropriate account of the distribution of health in different periods of an individual's life. Effective and appropriate decision-making concerning the allocation of health care resources therefore clearly requires the development of dyanamic measures of population health.

Measurement over time is always difficult, regardless of whether the focus is on poverty, inequality or health, etc. Although some significant progress has already been made in thinking about how to undertake chronic poverty measurement (that is, poverty over time), there is still no single compelling index or approach that effectively captures the dynamics of changes in the poverty profile at both the population and individual levels. For example, Jalan and Ravallion (2000) proposed an approach that takes the time mean of the squared poverty gap for household i as the measure of i's poverty profile over time and views the household-size weighted mean of the household poverty profiles (that is, the aggregate squared poverty gap) across the whole population as an aggregate measure of poverty for the population over time. However, their proposed measure is in fact insensitive to fluctuations in the distribution of the poverty profiles at both the individual and population levels. Foster (2007) develops a duration-adjusted Foster-Green-Thorbeck (FGT) measure, i.e. the mean of the α power of normalized gaps for all periods experienced by all chronically poor people. This measure is sensitive to changes in the distribution of income at the population level, but is not necessarily sensitive to changes over time in the poverty profile of individuals if this does not affect the actual distribution of income in the

population in each period of time. Calvo and Dercon (2007) explore the possibility of axiomatizing the measurement of poverty over time at the level of the individual, but they do not address the issue of aggregating individual measures to the level of the population as a whole.

Relatively little work has been done on the dynamic measurement of income inequality, which is very surprising given the richness of the literature on the static measurement of income inequality. In particular, to the best of our knowledge, there is no work that has been done using stochastic dominance to compare income inequality over time. The same story is generally true for the poverty measurement over time. Similarly, with the exception of work focusing on changes in life expectancy and quality adjusted life years[17], there has been little progress made in developing truly dynamic measures of population health.

Whereas we cannot expect to resolve all of the issues that need to be addressed in measuring population health over time, in this chapter we develop a reasonable candidate for a dynamic measure of population health, based on a set of axioms which we argue should be satisfied if one is to develop a satisfactory dynamic measure of population health. We first identify a number of axioms which we will argue should be satisfied by any useful measure of population health over time. Subsequently, we construct a measure which

[17] Although researchers have given attention to the way in which the QALY model aggregates information about individual health experience over a multi-year horizon (e.g., Richardson et al, 1996, Spencer and Robinson, 2007) they have focused mainly on examining whether the QALY scale remains valid when health states are chronic. In particular, they have been interested in whether or not the measure of an individual's health profile over time can be estimated by adding up the independently-rated constituent health states. Bleichrodt and Filko (2008) have tested the validity of QALYs to measure health over time under uncertainty, based on the concept of generalized marginality which deals with the preference relation on the independence of health in different periods. They conclude that although the QALY is additive over time,it does not incorporate the decision maker's preference structure, which seriously limits its usefulness in intertemporal policy-making contexts.

satisfies these properties. The approach taken here does not go as far as a full axiomatisation of the proposed dynamic measure of population health. We demonstrate that there is at least one measure - the Dynamic Mean Inequality Measure (DMIM) - which satisfies our proposed set of axioms; however, we do not prove that the only measure of population health which satisfies these axioms is the DMIM. Subsequently, with the help of CHNS longitudinal data, we illustrate the application of $DMIM$. The last section concludes.

3.2 A Critical Review of Axioms Frequently Used for the Measurement of Poverty and Income Inequality

In seeking to develop an appropriate - and operationally-useful - dynamic measure of population health, it is evident that much can be learned from the extensive literatures on income inequality and on poverty. Over the past few decades, considerable effort has been invested in axiomatizing the measurement of both income inequality and poverty, and consequently a careful review of some of the most commonly used axioms in these fields will provide valuable guidance in seeking to develop a meaningful measure of dynamic population health. The axiomatic approach to the measurement of poverty and income inequality is widely favoured by theorists, because it provides clear ethical foundations for the construction of measures of poverty or of income inequality. The measurement of population health exhibits some of the same -but not all - of the challenges encountered in appropriately measuring income inequality; this is also true when comparing the measure of population health and poverty. As with income inequality, the health of all individuals matters - rather than merely those whose level of health falls below some 'health poverty' line. As

with the analysis of absolute poverty, however, absolute individual health status truly matters: it is - at least in principle - an objectively meaningful statement to claim that health levels in one population are strictly greater than in another population. Consequently, an adequate measure of population health must find some way to balance distributional considerations with a concern for the absolute level of health attained by individuals in the community.

We first consider axiomatic approaches to the measurement of poverty. Although different authors typically adopt slightly different sets of axioms, the set of axioms proposed as an ethical foundation for the measurement of poverty typically include variants of the focus, monotonicity and transfer axioms. Let an income profile $x = \{x_1, x_2, ..., x_n\}$ describe the income x_i allocated to each individual i. Likewise, an income profile $y = \{y_1, y_2, ..., y_n\}$ describes the income y_i allocated to each individual i. The set of possible income profiles is denoted by D, and z is a scalar which denotes the poverty line, i.e., if the income of individual i is less than z, then individual i is considered to be poor. P is the measure of poverty.

Arguably the most commonly invoked axiom is the Focus Axiom: $P(x, z) = P(y, z)$, whenever $x \in D$ is obtained from $y \in D$ by an increment to a non-poor person. In interpretation, the focus axiom implis that a change in the distribution of income amongst the non-poor should not affect the measured level of poverty. In the context of the measurement of population health, it is not reasonable to consider a redistribution of health amongst healthy people: even if this were feasible - and, unlike income, health cannot be easily transferred from one individual to another - it is evident that if two individuals are both in

perfect health that it is not meaningful to undertake conceptual experiments in which health would be redistributed amongst perfectly healthy individuals. Whereas a variant on the focus axiom would perhaps be of interest if one were developing a measure of health poverty in the population, this is not useful in developing a dynamic measure of population health.

An almost equally frequently adopted axiom in the literature on poverty measurement is the **Monotonicity Axiom** (or Pareto principle): $P(x,z) < P(y,z)$ whenever $x \in D$ is obtained from $y \in D$ by a simple increment to a poor person. In interpretation, this axiom requires that a decrease (increase) in the income of a poor person should increase (decrease) the overall poverty level. Whereas this axiom initially appears attractive, it is evident that it is in conflict with egalitarian principles: in the eyes of egalitarians, if there is an initial situation in which income (or health) is equally distributed, then there is no social improvement (and arguably there may be a real social loss) if the income (or health) of just one individual is subsequently increased. A dynamic measure of population health which is sensitive to changes in the distribution of the health of individuals may therefore not always be consistent with the Monotonicity Axiom.

Another frequently invoked axiom in poverty measurement is the **Transfer Axiom** (Pigou-Dalton Principle): $P(x,z) > P(y,z)$ whenever $x \in D$ is obtained from $y \in D$ by a regressive transfer where the donor (and possibly the recipient) is poor. It requires that the poverty measure be sensitive to the redistribution of income between the poor and those who are still poorer. Although health cannot be easily redistributed between individuals, this axiom can be meaningfully re-interpreted when comparing alternative distributions of health across a given population, or across different populations. In the context of the

measurement of population health, when comparing two populations which are identical except with respect to the health of two individuals who are less healthy than the median individual, the population in which the least healthy individual has a lower level of health, and the more healthy individual has a higher level of health, is that which should be judged less healthy overall.

Turning now to the axioms used to underly measures of income inequality, it is perhaps not surprising that the **Pigou-Dalton Principle** is (suitably reinterpreted) an important feature of most of these measures. However, another important axiom is that **Symmetry Axiom** (or Principle of Anonmyity), which formalizes the notion that it is not the identity of the individuals to whom particular amounts of income are attributed that is important, but rather the distribution of income across the population. More precisely, if I denotes the measure of income inequality, defining the income profile $x = \{x_1, x_2, ..., x_n\}$ as above, and leting H denote the population size, we say that $I(x, H) = I(Px, H)$ for all permutation matrices P, where there is only one entry in each row and column having value equal to one, and all other entries are equal to zero. In interpretation, the symmetry axiom requires that the inequality measure is not affected by the order in which individuals are ranked.

Although there are a number of other axioms which are relatively often adopted in developing measures of income inequality, many are not of great relevance for the purposes of developing a dynamic measure of population health. However, one important exception is the **Decomposability Axiom:** $I(x_1, ..., x_G, H) = \sum_{g=1}^{G} w_g^G I(x_g, H^g) + I(\mu_1 e_1, ..., \mu_G e_G, H)$,where H^g denotes the number of individuals in subgroup g, x_g the income vector within group g, μ_g the mean income of group g , e_G the unit vector $(1, 1, 1..., 1)$,

and $w_g^G = \frac{\sum_{h=1}^{H^g} x^h}{\sum_{h=1}^{H} x^h}$ the ratio of the total income in group g to the total income in the population. This axiom requires that the measure of inequality be decomposable into the sum of within group inequality and between group inequality, and contrasts with the equivalent axiom in poverty measurement, which states that the overall measure of poverty can be decomposed into that of subgroups according to certain characteristics, but does not take account of the "between group" effect.

3.3 Appropriate Axioms for the Dynamic Measurement of Population Health

Although the objective of this analysis is not to develop a full axiomatisation of a dynamic measure of population health, it is evident that candidate dynamic measures will need to satisfy a number of key axioms for them to be judged worthy of more careful study and analysis. Consequently, in this section, we seek to identify a number of axioms that appear to be particularly compelling when considering the measurement of population health. As before, we tackle this task by considering a series of simple examples which highlight some key issues arising in undertaking comparisons of the health of different populations over a given period of time. This exercise draws on the five criteria we discussed in Chapter 1, appropriately enriched to reflect the specific issues which arise when seeking to measure population health in the dynamic context.

It is useful to introduce some notation. Let $X_t = (x_{1,t}, x_{2,t}, ..., x_{n,t})$ be a vector describing the health of each member of a given population at time t, and $x_i = (x_{i,t=1}, x_{i,t=2}, ..., x_{i,t=T})$ describe the profile of individual i's health status over time, from period $t = 1$ to period

$t = T$. The health of the population over time is then described by a $T \times n$ matrix X with each entry $x_{i,t}$. In any period when the individual is dead, $x_{i,t} = 0$.

$$X = \begin{pmatrix} x_{1,t=1}, x_{2,t=1}, ..., x_{n,t=1} \\ x_{1,t=2}, x_{2,t=2}, ..., x_{n,t=2} \\ ... \\ x_{1,t=T}, x_{2,t=T}, ..., x_{n,t=T} \end{pmatrix}$$

We denote the dynamic measure of population health as $H(X)$

3.3.1 Monotonicity Axiom (Weak Pareto Principle)

In comparing the health of two populations A and B over time, if the only difference between the two populations is that one individual in population A experiences higher health in each period than the equivalently-ranked individual in population B, it is natural to conclude that $H(A) > H(B)$. This is the same ethical view as underlies the monotonicity axiom in the case of the measurement of poverty: more relatively healthy individuals should increase population health, just as fewer poor people (or a decrease in poverty amongst the poor) reduces poverty. In the context of dynamic population health measurement, a monotonicity axiom could be expressed as:

If

$$X = \begin{pmatrix} x_{1,t=1}, x_{2,t=1}, ..., x_{n,t=1} \\ x_{1,t=2}, x_{2,t=2}, ..., x_{n,t=2} \\ ...x_{i,t=j}... \\ x_{1,t=T}, x_{2,t=T}, ..., x_{n,t=T} \end{pmatrix}$$

and

$$X' = \begin{pmatrix} x_{1,t=1}, x_{2,t=1}, ..., x_{n,t=1} \\ x_{1,t=2}, x_{2,t=2}, ..., x_{n,t=2} \\ ...x_{i,t=j} + \Delta x... \\ x_{1,t=T}, x_{2,t=T}, ..., x_{n,t=T} \end{pmatrix}$$

then $H(X') > H(X)$ for $\Delta x > 0$.

In contrast, of course, axiomatic foundations for the measurement of income inequality typically do not include an axiom of this sort, because inequality is viewed as undesirable: if in the initial situation, income is equally distributed across the entire population, and then there is an alternative in which all individuals have the same income except for one person who experiences an increase in income, then there is greater income inequality in the second situation as compared to the first, even though average income is higher.

In the context of the dynamic measurement of population health, there would seem to need to be some sort of balance between, on the one hand, recognizing the importance of improvements in the health of individuals and, on the other hand, putting appropriate weight on the actual distribution of health in the population. Just as one wishes measures of poverty to be sensitive to improvements in the welfare of the poor, but not accord particular importance to improvements in the welfare of the non-poor, it seems appropriate to suggest that a compelling measure of dynamic population health should be sensitive to improvements in the health of the relatively unhealthy (i.e., those whose health level lies below the median level), but not require that the measure be sensitive to increases in the health level of those who are already in relatively good health. This can be expressed as:

If

$$X = \begin{pmatrix} x_{1,t=1}, x_{2,t=1}, \ldots, x_{n,t=1} \\ x_{1,t=2}, x_{2,t=2}, \ldots, x_{n,t=2} \\ \ldots x_{i,t=j} \ldots \\ x_{1,t=T}, x_{2,t=T}, \ldots, x_{n,t=T} \end{pmatrix}$$

and

$$X' = \begin{pmatrix} x_{1,t=1}, x_{2,t=1}, \ldots, x_{n,t=1} \\ x_{1,t=2}, x_{2,t=2}, \ldots, x_{n,t=2} \\ \ldots x_{i,t=j} + \Delta x \ldots \\ x_{1,t=T}, x_{2,t=T}, \ldots, x_{n,t=T} \end{pmatrix}$$

and $\widetilde{H}_i <$ median health, then $H(X') > H(X)$ for $\Delta x > 0$.

3.3.2 Replication invariance Axiom

Consider two populations of different overall size but which share the same demographic structure and in which the same proportion of the population experiences a given health status. In this case, the measure of population health should be independent of the size of the population. Thus, if

$$X = \begin{pmatrix} x_{1,t=1}, x_{2,t=1}, \ldots, x_{n,t=1} \\ x_{1,t=2}, x_{2,t=2}, \ldots, x_{n,t=2} \\ \ldots x_{i,t=j} \cdots \\ x_{1,t=T}, x_{2,t=T}, \ldots, x_{n,t=T} \end{pmatrix}$$

and

$$X' = \underbrace{X + \ldots + X}_{n} = nX$$

then $H(X') = H(X)$,for any integer n.

If identical populations are merged, then the population health is unchanged.

3.3.3 Weak anonymity Axiom

The aggregation in $H(X)$ is based on individual's overall health status during T periods. As noted earlier, health is non-transferrable between individuals. Consequently, it is particularly important that the dynamic measure of population health be sensitive to changes in the health status of individuals.

For example, consider the populations A, $A1$ and $A2$:

Population A					
period/individual	1	2	3	4	5
1	0.3	0.4	0.5	0.6	0.7
2	0.4	0.5	0.5	0.6	0.7
3	0.5	0.6	0.5	0.6	0.7
Population A1					
period/individual	1	2	3	4	5
1	0.4	0.3	0.5	0.6	0.7
2	0.5	0.4	0.5	0.6	0.7
3	0.6	0.5	0.5	0.6	0.7
Population A2					
period/individual	1	2	3	4	5
1	0.5	0.4	0.5	0.6	0.7
2	0.4	0.5	0.5	0.6	0.7
3	0.3	0.6	0.5	0.6	0.7
Population B					
period/individual	1	2	3	4	5
1	0.3	0.4	0.5	0.6	0.7
2	0.4	0.5	0.5	0.6	0.7
3	0.5	0.5	0.6	0.6	0.7

Comparing population A with population $A1$, the only difference is that the health experience of individuals 1 and 2 is interchanged. Since it is the distribution of health experience over individual lifetimes that should matter, and not the identification of the individuals who experience a particular sequence of health states, then by weak anonymity these two populations should have same measures. Also, comparing population A with population $A2$, the difference only exists in individual 1 where his health experience of period 1 is switched with that of period 3. Then the health of these two populations should still have same measures, by the assumption that it is only the fluctuation of health experience rather than the order of health experience that matters.

Note that if one takes the view that the order in which individuals experience different health states is unimportant, then this implies that health states are not discounted.

Clearly, this is a strong normative assumption, although supported by a few empirical studies.[18] From the point of view of public policy making, the introduction of discounting introduces awkward ethical dilemmas as it essentially implies that the poor health of individuals today is more of a problem than the poor health of citizens tomorrow. Moreover, from a purely practical point of view, there simply does not exist an intertemporal Health Utility Index which can be used to differentiate between alternative sequences of different health states for individuals: HUI is effectively a static measure. And although at least some alternative measures of individual health outcomes (such as the DALY) incorporate measures of discounting, they do so in what is an admittedly arbitrary fashion; there does not exist convincing evidence to provide guidance on choosing the discount rate (Murray and Acharya 2002). [19] Consequently, by setting $\rho = 1$, individuals who experience greater fluctuations in their health states are found to be less healthy than individuals with the same average health experience (but less fluctuation), but there is no actual discounting of future health states. And for this to work, a constant age distribution of the population is also necessary. However, if demographic factors and demographic changes are of interest to the analysis, then cohorts could be weighted to reflect the relative importance of the targeted

[18] Redelmeier and Heller (1993) observed zero discount rate in 62% of cases in a study of time preference rate over acute health states, and Dolan and Gudex (1995) found a zero median discount rate across six EQ-5D states of health.

[19] Hammitt (1993) argues that one way of answering the question of whether or not it is appropriate to discount a health increment is to ask whether the benefit of the increment in health can be monetized. As measures of individual health (such as the HUI or QALY) are developed using questionnaire-based methodologies, it is not clear how to compare them at different ages. Anand and Hansen (1997) have strongly questioned the appropriateness of discounting DALYs for future periods by presenting a counter-example which demonstrates that, with a discount rate of 3%, it is preferrable to save the life of a 20-year old person rather than an infant, which doesn't accord with general intuition.

age groups. That would be an interesting extension of this current study, however, it will not affect the assumption of no discounting of future health states.

Comparing population A with population B, where in each period the individuals' health is ordered from least to best, the population health of these populations should differ, because in population A the health levels experienced by individual 2 (respectively, individual 3) are between 0.4 and 0.6, (respectively, 0.5) whereas in B this individual experiences health levels between 0.4 and 0.5 (respectively, 0.5 and 0.6).

As compared to the static setting, where the measured level of population health is a function only of the distribution of individual health levels, in dynamic settings it seems reasonable to require that the measure of population health not care about which individual experiences a particular series of states of health (that is, anonymity should be maintained with respect to the vector of individual health levels), but anonymity should not be imposed at the level of each time period, that is, interchanging the experience of health at time t between two individuals, and therefore altering their vector of health states which each individual experiences over their lifetime, should lead to different levels of measured population health. More precisely, $H(X) = H(XP) = H(PX)$ for any permutation matrx P, also $H(X) = H(\cup X_t P)$, but $H(X) \neq H(\cup X_t P_t)$ if $P_i \neq P_j$.

The weak anonymity axiom plays a key role in distinguishing a meaningful dynamic measure of population health from one which merely captures the cross-sectional trend of the health of the population at every period. This would be the case, for example, if the measure were based on a sequence of static measures of population health, e.g., calculating the MIM for the population in each period, and then subsequently taking the

MIM of this sequence. A trend-based approach can be viewed as macro approach, whereas weak anonymity axiom as developed here is a micro approach. But whereas a trend-based approach provides useful information on changes in the profile of population health over time, it fails to reveal the full picture of the dynamics of population health, because it does not capture the fluctuation in each individual's health experience over time.

3.3.4 Transfer axiom

It is again important to stress that health is not transferrable between individuals, and therefore when proposing a variation of the transfer axiom to the measurement of population health, the exercise of comparing the different populations should be seen as a 'thought experiment'. Specifically, in comparing two otherwise identical populations, if it is the case that in one population a relatively healthy individual is worse off, and there is an equal and correponding increase in the health state of a relatively less healthy individual or, alternatively, if there is an equal and corresponding increase in the health of this same individual but in periods when this person is relatively unhealthy, then the measured health of the second population should exceed that of the first. In effect, this is equivalent to requiring that the measure of population health be a concave function of measures of individual health. This principle is illustrated by the examples of population A, $A3$ and $A4$.

Population A					
period/individual	1	2	3	4	5
1	0.3	0.4	0.5	0.6	0.7
2	0.4	0.5	0.5	0.6	0.7
3	0.5	0.6	0.5	0.6	0.7
Population A3					
period/individual	1	2	3	4	5
1	0.35	0.4	0.5	0.6	0.7
2	0.4	0.5	0.5	0.6	0.7
3	0.45	0.6	0.5	0.6	0.7
Population A4					
period/individual	1	2	3	4	5
1	0.35	0.4	0.5	0.6	0.7
2	0.4	0.5	0.5	0.6	0.7
3	0.5	0.55	0.5	0.6	0.7

In this example, in population $A3$ individual 1 experiences better health in period 1, and worse health in period 3, than in population A. Requiring that the measure of population health be concave in the health experience of individuals over their lifetime then implies that population $A3$ is healthier than population A.

Next, if we compare populations $A3$ and $A4$ we observe that in the second population a relatively healthy individual experiences a slight reduction in health, and there is an equal increase in the health of a relatively unhealthy individual. As before, if the measure of population health is concave in the health experience of individual citizens, population $A4$ should be considered healthier than population $A3$. More precisely, suppose that Q is a square, biostochastic matrix, with the property that each entry is nonnegative, and that each of its rows and columns sums to 1. Then $H(QX) > H(X)$.

3.3.5 Decomposability axiom

When the population can be divided into subgroups, either by source or type of disease, or by some other demographic characteristic of the population, it is of interest to know whether the measure of the health of the population can be aggregated from information about the subgroups. From a practical point of view, developing and evaluating the impact of targeted health policies is more straightforward if the measure of population health can be decomposed into measures of the health of subpopulations. Consequently, suppose that s_i is subgroup i of the population, that $\overline{s_i} = (\frac{1}{n_i T}\sum_{i=1}^{n_i}\sum_{t=1}^{T} h_{it})$denotes the average health level of subgroup i, that ϕ_i is the weight assigned to subgroup i in determining the contribution of this subgroup to the health of the population, and ε is the residual term. Then if the initial distribution of health states across individuals in the population is described by the matrix $H(X)$ where

$$X = \begin{pmatrix} x_{1,t=1}, x_{2,t=1}, \ldots, x_{n,t=1} \\ x_{1,t=2}, x_{2,t=2}, \ldots, x_{n,t=2} \\ \ldots x_{i,t=j} \cdots \\ x_{1,t=T}, x_{2,t=T}, \ldots, x_{n,t=T} \\ \underbrace{\qquad\qquad}_{s_1} \cdots \underbrace{\qquad\qquad}_{s_s} \end{pmatrix}$$

then the measure of population health satisfies the decomposability axiom if $H(X) = \sum \phi_i H(s_i) + H(\sum \overline{s_i}) + \varepsilon$

For the reasons discussed above, it is our view that a satisfactory dynamic measure of population health should be consistent with these axioms. However, it is also evident that it is important to relate these axioms to the five criteria set forth to evaluate the appropriateness of a measure of population health in Chapter 1. The issue of completeness of the ranking is implicitly considered: the dynamic version of the MIM proposed be-

low inherits the property of completeness from the MIM measure. The monotonicity in the second criterion is reflected in the monotonicity axiom, appropriately extended here to accommodate the dynamic setting. The convexity in the third crierion and the tradeoff between health inequality and average health in the fifth criterion is crystalized here in the transfer axiom. The fourth criterion is implicitly embedded in the replication and weak anonymity axioms.

Below, we propose a dynamic version of the MIM measure - the DMIM - and we show that this measure will indeed satisfy these axioms.

3.4 Dynamic version of MIM ($DMIM$)

The objective in developing a dynamic version of MIM ($DMIM$) is to provide greater insight into how different distributions of population health can be compared, when these distributions provide data about the health of individuals over time. The measurement proposed views aggregate population health as being constructed in two stages. The first is designed to capture differences in the health experience of individuals, and puts weight on both the average level of health as well as variation in the individual's health experience across time. Then, in the second stage, the MIM is computed for the population as a whole, using the measured levels of lifetime individual health as the data. More precisely, if

$$X = \begin{pmatrix} x_{1,t=1}, x_{2,t=1}, \ldots, x_{n,t=1} \\ x_{1,t=2}, x_{2,t=2}, \ldots, x_{n,t=2} \\ \cdots \\ x_{1,t=T}, x_{2,t=T}, \ldots, x_{n,t=T} \end{pmatrix}$$

then in stage one, for each individual i we calculate

$$\begin{aligned} MIM_i &= \overline{g}_i + h_i I_i = \widetilde{G}_i, \quad \widetilde{H}_i = 1 - \widetilde{G}_i \\ \text{where } h_i &= \frac{\sum_{t=1}^{T} h_{i,t}}{T}, \overline{g}_i = 1 - h_i \text{ and } I_i = I(h_{i,1}, ..., h_{i,T}). \end{aligned}$$

Then, in stage two, we apply the MIM measure to the population, i.e.,

$$DMIM(X) = [\alpha \widetilde{G}^{\rho} + \beta(\widetilde{H}\widetilde{I})^{\rho}]^{1/\rho}$$

where $\widetilde{G} = \frac{\sum_{i=1}^{n} \widetilde{G}_i}{n}$, $\widetilde{H} = 1 - \frac{\sum_{i=1}^{n} \widetilde{G}_i}{n}$, $\widetilde{I} = I(\widetilde{H}_{1,...}, \widetilde{H}_n)$, and $\rho \leq 1$.

The DMIM measure is relatively straightforward to calculate. However, it remains to be shown that it is also consistent with the axioms we have identified above which we have argued should be satisfied by any appropriate dynamic measure of population health.

3.4.1 Pareto sensitivity

Proposition 6 *For a given intertemporal health allocation across individuals,an increase in the health experience of any individual whose health status lies below the median health status for the population, at any period of time t, reduces the $DMIM$. This implies that population health is increased. Consequently, DMIM satisfies the Weak Pareto Axiom.*

Proof. Suppose there is an increment Δh in health to an individual i at period t.Then by Lemma 4 (Chapter 2), MIM_i is monotonic when $\rho = 1$. Moreover, since $\widetilde{G}_i$ decreases for all individuals who have experienced a health increment, it must be the case that $\widetilde{G}' < \widetilde{G}$.

Turning to the second stage of computing $DMIM$, we notice that the increase in the health of any invididual may either increase or decrease the health inequality of the population, i.e., $I' > I$ or $I' < I$.

When $\rho = 1$, and again following the proof of Lemma 4 in Chapter 2, the change in health inequality is dominated by the change in average health. It is then immediate that the increment to the individual's health will increase the health of the population.

It remains to show that when $\rho < 1$,then this always results in a smaller $DMIM$ value if the increment of health accrues to an individual ranked below the median position after their increase in health. Assume that an increase Δg accrues to individual $\bar{j}$, where j is the rank of the individual's health in the population, ordered from lowest to highest. Suppose, initially, that the increment of health is small enough so that it does not affect the rankings of individuals' health. In this case, the change in health inequality can be expressed as

$$\Delta I = \frac{2\Delta g(\sum(\bar{j} - i)\widetilde{H}_i)}{n\sum \widetilde{H}_i(\sum \widetilde{H}_i + \Delta g)}.$$

Denote the initial distribution of health in the population as A, and the final distribution of health in the population as B. Then, by definition, $DMIM_A = [\widetilde{G}_A^{\rho} + (\widetilde{H}_A\widetilde{I}_A)^{\rho}]^{1/\rho}$ and consequently

$$\begin{aligned} DMIM_B &= [(\widetilde{G}_A - \frac{\Delta g}{n})^{\rho} + (\widetilde{H}_A + \frac{\Delta g}{n})^{\rho}(\widetilde{I}_A + \Delta I)^{\rho}]^{1/\rho} \\ &= [(\widetilde{G}_A - \frac{\Delta g}{n})^{\rho} + (\widetilde{H}_A\widetilde{I}_A + \frac{\Delta g}{n}\frac{2j - n - 1}{n})^{\rho}]^{1/\rho}, \end{aligned}$$

Then, taking a first order approximation,

$$\begin{aligned} DMIM_B &= [\widetilde{G}_A^{\rho} - \rho\widetilde{G}_A^{\rho-1}\frac{\Delta g}{n} + (\widetilde{H}_A\widetilde{I}_A)^{\rho} + \rho(\widetilde{H}_A\widetilde{I}_A)^{\rho-1}\frac{\Delta g}{n}\frac{2j-n-1}{n}]^{1/\rho} \\ &= [\widetilde{G}_A^{\rho} + (\widetilde{H}_A\widetilde{I}_A)^{\rho} - \rho\overline{g}_A^{\rho-1}\frac{\Delta g}{n} + \rho(\bar{h}_A I_A)^{\rho-1}\frac{\Delta g}{n}\frac{2j-n-1}{n}]^{1/\rho}. \end{aligned}$$

Notice that if individual j is ranked below the median, i.e., $j \leq \frac{n}{2}$ when n is even, or $j \leq \frac{n+1}{2}$ when n is odd, then $\frac{2j-n-1}{n} \leq 0$, and so when $0 < \rho \leqslant 1$ it must be true that

$$-\rho\overline{g}_A^{\rho-1}\frac{\Delta g}{n} + \rho(\bar{h}_A I_A)^{\rho-1}\frac{\Delta g}{n}\frac{2j-n-1}{n} \leq 0$$

which can be re-expressed as

$$\widetilde{G}_A^{\rho} + (\widetilde{H}_A\widetilde{I}_A)^{\rho} - \rho\overline{g}_A^{\rho-1}\frac{\Delta g}{n} + \rho(\bar{h}_A I_A)^{\rho-1}\frac{\Delta g}{n}\frac{2j-n-1}{n} \leq \widetilde{G}_A^{\rho} + (\widetilde{H}_A\widetilde{I}_A)^{\rho}.$$

Recalling that x^{θ} is an increasing function when $\theta > 0$,it follows immediately that $DMIM_B \leq DMIM_A$.

Alternatively, if $\rho < 0$,

$$-\rho\overline{g}_A^{\rho-1}\frac{\Delta g}{n} + \rho(\bar{h}_A I_A)^{\rho-1}\frac{\Delta g}{n}\frac{2j-n-1}{n} \geq 0$$

which can be re-expressed as

$$\widetilde{G}_A^{\rho} + (\widetilde{H}_A\widetilde{I}_A)^{\rho} - \rho\overline{g}_A^{\rho-1}\frac{\Delta g}{n} + \rho(\bar{h}_A I_A)^{\rho-1}\frac{\Delta g}{n}\frac{2j-n-1}{n} \geq \widetilde{G}_A^{\rho} + (\widetilde{H}_A\widetilde{I}_A)^{\rho}.$$

Once again, because x^{θ} is a decreasing function when $\theta < 0$, it again follows immediately that $DMIM_B \leq DMIM_A$.

Next consider the more general case: suppose that the increment of health to individual j changes that person's rank in the population to $J + K$, and therefore also affects

the ranking of individuals whose health levels were originally ranked from $J+1$ to $J+K$. In this case, the change in health inequality can be expressed as:

$$\Delta I = \frac{2[\Delta g \sum(\bar{j}+K-i)\widetilde{H}_i - \sum_{i=J+1}^{J+K}(\widetilde{H}_i - \widetilde{H}_J)\sum \widetilde{H}_i]}{n\sum \widetilde{H}_i(\sum \widetilde{H}_i + \Delta g)}.$$

Once again, denote the original distribution of health states in the population as A and the final distribution of health states in the population as B. By definition, $DMIM_A = [\widetilde{G}_A^{\rho} + (\widetilde{H}_A\widetilde{I}_A)^{\rho}]^{1/\rho}$. This implies that

$$\begin{aligned} DMIM_B &= [(\widetilde{G}_A - \frac{\Delta g}{n})^{\rho} + (\widetilde{H}_A + \frac{\Delta g}{n})^{\rho}(\widetilde{I}_A + \Delta I)^{\rho}]^{1/\rho} \\ &= [(\widetilde{G}_A - \frac{\Delta g}{n})^{\rho} + (\widetilde{H}_A\widetilde{I}_A + \frac{\Delta g(2(j+k)-n-1) - 2\sum_{i=J+1}^{J+K}(\widetilde{H}_i - \widetilde{H}_J)}{n^2})^{\rho}]^{1/\rho}, \end{aligned}$$

which, by first order approximation, can be re-expressed as:

$$\begin{aligned} DMIM_B &= [\widetilde{G}_A^{\rho} - \rho\widetilde{G}_A^{\rho-1}\frac{\Delta g}{n} + (\widetilde{H}_A\widetilde{I}_A)^{\rho} \\ &\quad +\rho(\widetilde{H}_A\widetilde{I}_A)^{\rho-1}\frac{\Delta g(2(j+k)-n-1) - 2\sum_{i=J+1}^{J+K}(\widetilde{H}_i - \widetilde{H}_J)}{n^2}]^{1/\rho} \\ &= [\widetilde{G}_A^{\rho} + (\widetilde{H}_A\widetilde{I}_A)^{\rho} - \rho\overline{g}_A^{\rho-1}\frac{\Delta g}{n} \\ &\quad +\rho(\bar{h}_A I_A)^{\rho-1}\frac{\Delta g(2(j+k)-n-1) - 2\sum_{i=J+1}^{J+K}(\widetilde{H}_i - \widetilde{H}_J)}{n^2}]^{1/\rho} \end{aligned}$$

where $\sum_{i=J+1}^{J+K}(\widetilde{H}_i - \widetilde{H}_J)$ is the sum of gaps between individual J and individuals whose ranks are affected by the increase in individual's health. Notice that $0 \leq \sum_{i=J+1}^{J+K}(\widetilde{H}_i - \widetilde{H}_J) \leq K\Delta g$, depending on the original distribution of health in the population. Consider, then, that $\sum_{i=J+1}^{J+K}(\widetilde{H}_i - \widetilde{H}_J) = K\Delta g$, and assume that the increment of health arises to an individual whose health level is less than the median level of health, i.e., $j \leq \frac{n}{2}$ when

n is even, or $j \leq \frac{n+1}{2}$ when n is odd. Then, as before, this implies that $\frac{2j-n-1}{n} \leq 0$ and therefore for all $\rho \leq 1, \rho \neq 0$, $DMIM_B \leq DMIM_A$.

Similarly, if $\sum_{i=J+1}^{J+K}(\widetilde{H}_i - \widetilde{H}_J) = 0$, and given our requirement that individual j still be below the median after experiencing an improvement in their state of health, we know that if $(j+k) \leq \frac{n}{2}$ when n is even, or $(j+k) \leq \frac{n+1}{2}$ when n is odd. Therefore, it must be true that $\frac{2(j+k)-n-1}{n} \leq 0$ for all $\rho \leq 1 (\rho \neq 0)$, and consequently $DMIM_B \leq DMIM_A$. ■

This proposition establishes that the DMIM measure satisfies the Weak Pareto principle: increments in health which accrue to individuals whose rank in the health distribution is lower than the median - and whose ranking does not surpass the median ranking as a result of their improved health status - result in measured improvements to the health of the population.

Corollary 7 *Let $\rho = 1$. If health is fully equally distributed, so that $I = 0$, then this reduces $DMIM$ and therefore population health increases.*

Proof. In the absence of health inequality, it is necessarily the case that $DMIM_A = \widetilde{G}_A$, since $\widetilde{I}_A = 0$. Consequently, if any individual i experiences an improvement in their health, this increase health inequality:

$$\Delta I = \frac{(n+1)(\sum \widetilde{H}_i + \Delta g) - 2(\sum (n-i+1)\widetilde{H}_i + \Delta g)}{n(\sum \widetilde{H}_i + \Delta g)}$$

which means that

$$\begin{aligned} DMIM_B &= [(\widetilde{G}_A - \frac{\Delta g}{n})^\rho + (\widetilde{H}_A + \frac{\Delta g}{n})^\rho (\Delta I)^\rho]^{1/\rho} \\ &= [(\widetilde{G}_A - \frac{\Delta g}{n})^\rho + (\frac{\Delta g}{n}\frac{n-1}{n})^\rho]^{1/\rho}. \end{aligned}$$

However, when $\rho = 1$,

$$DMIM_B = \widetilde{G}_A - \frac{1}{n^2}\Delta g$$

so $DMIM_B < DMIM_A$ and therefore $DMIM$ satisfies the strong Pareto principle. ■

In contrast, if $\rho < 1$,then if in the initial situation health is equally distributed across the population, and one individual experiences a positive increase in their health, this may actually lower the measured health of the population. In effect, when the average health gap and the size of the population are both sufficiently large, and when the aversion to inequality is strong enough, an increase in the health of any one individual does not result in a significant change to the average health of the population, but it may lead to a significant increase in the measured level of health inequality, leading to a lower measure of population health.

3.4.2 Replication invariance

As discussed above, it is important that the measure of population health depend only on the distribution of health across individuals in the population, rather than to the number of individuals experiencing given health states.

Proposition 8 *Suppose that there are two populations A and B consisting of the same number of individuals, and with the same distribution of individual health states across these individuals. Then $DMIM(A) = DMIM(B) = DMIM(A \cup B)$ for any $\rho \leq 1 (\rho \neq 0)$.*

Proof. This follows immediately from the fact that the average health gap in the merged population is the same as in the constituent populations and that the Gini coefficient is sensitive only to the relative proportions of the total population in each health state, and not the number of individuals. ■

3.4.3 Weak anonymity

Proposition 9 *For a given health allocation X among n individuals, the measure of population health $DMIM$ is invariant to the positions of individuals in the allocation, as long as the profile of health experiences of each individual does not change.*

Proof. Trivial. ■

Note that, by construction, the $DMIM$ generates the same measured level of population health for any populations in which the same proportion of individuals experience the same health states. However, because $\rho = 1$ at stage one, the order in which they experience these health states is irrelevant. In contrast, $DMIM$ generates different measures of population health if the set of health states experienced by individuals differs, *even if* the actual set of health states experienced by the population as a whole remains the same.

3.4.4 Transfer sensitivity

As discussed above, there are two distinct notions of transfer sensitivity, and it is desirable that a dynamic measure of population health be sensitive to both sorts of transfers. The first situation involves comparisons of otherwise identical populations, but in which the health experience of a single individual is changed to improve that individual's health in

relatively poor states, and diminished in relatively healthy states. The second situation involves comparisons which 'shift' health from a relatively healthy to a relatively less healthy individual.

Proposition 10 *Consider two populations A, A' which differ only with respect to the health of individual a, but where the average health of individual a is the same in both populations. Let $x_{a,t}(A) \geq x_{a,t}(A')$ for all t such that $x_{a,t}(A) \leq \sum_{t=0}^{T} x_{a,t}(A)/\|T\|$, and $x_{a,t}(A) \leq x_{a,t}(A')$ for all t such that $x_{a,t}(A) \geq \sum_{t=0}^{T} x_{a,t}(A)/\|T\|$. Then there exists $\rho^* < 1$ such that $DMIM_A \leq DMIM_{A'}$ for any $\rho \in (\rho^*, 1]$.*

Proof. By construction, we are considering a mean-preserving transfer across health periods, and therefore the measure of individual health after the change depends only on the change of health inequality aross health periods for individual a, i.e I_a. note that

$$I_a = 1 + \frac{1}{n} - \frac{2\sum(n-i+1)\widetilde{H}_i}{n\sum \widetilde{H}_i}$$

and therefore, recalling that $\rho = 1$ in the first stage,

$$\begin{aligned} I'_a &= 1 + \frac{1}{n} - \frac{2[\sum_{i\neq j,k}(n-i+1)\widetilde{H}_i + (n-j+1)(\widetilde{H}_j - \Delta g) + (n-k+1)(\widetilde{H}_k + \Delta g)]}{n\sum \widetilde{H}_i} \\ &= 1 + \frac{1}{n} - \frac{2[\sum(n-i+1)\widetilde{H}_i] + 2(j-k)\Delta g}{n\sum \widetilde{H}_i} \end{aligned}$$

where $j \geq k$. In short

$$dI_a = [\frac{2dg}{n\sum \widetilde{H}_i}](j-k) \leq 0.$$

Consequently, individual a has better overall health. It follows directly that since the average health of the population is higher in population A, then by Proposition 3 (Chapter 2), we know that there exists a $\rho^* < 1$ (which equalizes the change in average health and

health inequality) such that for all $\rho > \rho^*$ it is necessarily true that $DMIM_A \leq DMIM_{A'}$.

■

Proposition 11 *Consider the distribution of health across two populations, A and B, which are identical except with respect to the health of two individuals. Suppose that in population B there is a reduction in the state of health of one individual during the kth most healthy period in this individual's life, and there is a corresponding increase in the health experience of a less healthy individual in population B in the jth most healthy period of this individual's life, where $j < k$. Then there exists $\rho^* < 1$ such that $DMIM_B \leq DMIM_A$ for any $\rho \in (\rho^*, 1]$.*

Proof. Consider two individuals a and b such that, in stage one, $MIM_a = \widetilde{G}_a < MIM_b = \widetilde{G}_b$, i.e., individual a is healthier than individual b. Now suppose that there is a 'transfer' of δ units of health from $h_{a,t}$ to $h_{b,t'}$, where $h_{a,t} - \delta \geq h_{b,t'} + \delta$. We would like to prove that $DMIM > DMIM'$. Recall that in stage one $\rho = 1$ and therefore by construction,

$$MIM_a = \overline{g}_a + h_a I_a, MIM_b = \overline{g}_b + h_b I_b,$$

which implies that after the transfer

$$\begin{aligned} MIM'_a &= (\overline{g}_a + \frac{\delta}{T}) + (h_a - \frac{\delta}{T})I'_a \\ MIM'_b &= (\overline{g}_b - \frac{\delta}{T}) + (h_b + \frac{\delta}{T})I'_b. \end{aligned}$$

Denoting $\Delta MIM_a = MIM'_a - MIM_a$and $\Delta MIM_b = MIM'_b - MIM_b$, transfer sensitivity requires $\mid \Delta MIM_b \mid > \mid \Delta MIM_a \mid$. But this requirement is equivalent to

$$MIM_b - MIM'_b > MIM'_a - MIM_a$$

i.e.,

$$\begin{aligned}
& MIM_a + MIM_b > MIM'_a + MIM'_b \\
\Rightarrow\ & h_a I_a + h_b I_b > (h_a - \frac{\delta}{T})I'_a + (h_b + \frac{\delta}{T})I'_b \\
\Rightarrow\ & h_a I_a + h_b I_b > h_a I_a - \frac{\delta}{T} I_a + h_a \Delta I_a + \frac{\delta}{T}\Delta I_a + h_b I_b + \frac{\delta}{T} I_b + h_b \Delta I_b + \frac{\delta}{T}\Delta I_b \\
\Rightarrow\ & -\frac{\delta}{T} I_a + h_a \Delta I_a + \frac{\delta}{T}\Delta I_a + \frac{\delta}{T} I_b + h_b \Delta I_b + \frac{\delta}{T}\Delta I_b < 0 \\
\Rightarrow\ & \Delta I_a (h_a - \frac{\delta}{T}) - \frac{\delta}{T} I_a + \Delta I_b (h_b + \frac{\delta}{T}) + \frac{\delta}{T} I_b < 0.
\end{aligned}$$

Since

$$\begin{aligned}
\Delta I_a &= \frac{2\delta(\sum (i - \bar{k})\widetilde{H}_{a,i})}{T \sum \widetilde{H}_{a,i} (\sum \widetilde{H}_{a,i} - \delta)} \\
\Delta I_b &= \frac{2\delta(\sum (\bar{j} - i)\widetilde{H}_{b,i})}{T \sum \widetilde{H}_{b,i} (\sum \widetilde{H}_{b,i} + \delta)}
\end{aligned}$$

we have

$$\frac{\delta}{T}\frac{2j - T - 1}{T} + \frac{\delta}{T}\frac{T - 2k + 1}{T} < 0$$

Therefore, it requires that $j < k$, that is, if the health experience over the lifetime of the two individuals are rank-ordered, the transfer must be from a higher-ranked period in the life of the healthier individual to a (relatively) lower-ranked period in the life of the less-healthy individual. At the second stage, by Proposition 3, Chapter 2, we know that there exists a $\rho^* \leq 1$ such that for all $\rho > \rho^*$, $DMIM_B \leq DMIM_A$.In summary, when there

is a transfer from a healthier period of an individual to a less healthy period of another individual, the $DMIM$ satisfies transfer sensitivity. ■

3.4.5 Decomposability

It remains to investigate whether or not the DMIM measure is consistent with the principles of decomposability outlined above.

Proposition 12 *Suppose that* $\rho = 1$ *and that the population can be viewed as consisting of* L *subgroups. Then* $DMIM$ *can be decomposed into an expression which measures overall population health as an aggregation of within group components and between group components for each of the* L *subgroups, minus an adjusted average health gap.*

Proof. By definition, when $\rho = 1$, $DMIM = [\widetilde{G} + (\widetilde{H}\widetilde{I})]$. Suppose that there are L subgroups in the population which do not have overlapping membership. Then $DMIM = \widetilde{G} + \widetilde{H}\widetilde{I}$. We now show that this measure can be expressed as $DMIM = \sum_{i=1}^{L} \phi_i^2 DMIM_i + \overline{DMIM} - \sum_{i=1}^{L} \phi_i^2 G_i$.

Observe that, if there are L subgroups, then $I = \sum_{j=1}^{N}[\frac{\overline{\delta}_j}{\widetilde{H}N}] = \sum_{i=1}^{L}\sum_{k=1}^{q_i}[\frac{\phi_i\overline{\delta}_{k,i}+\widetilde{\delta}_{k,i}}{\widetilde{H}N}] = \sum_{i=1}^{L}[\phi_i^2\frac{H_i}{\widetilde{H}}\frac{\sum_{k=1}^{q_i}\overline{\delta}_k}{H_i q_i}] + \sum_{i=1}^{L}\sum_{k=1}^{q_i}\frac{\widetilde{\delta}_k}{\widetilde{H}N}$, where $\overline{\delta}_k = \frac{\sum_{j=1}^{N}(\widetilde{H}_j - \widetilde{H}_k)_+}{N} = \phi_i\overline{\delta}_{k,j} + \widetilde{\delta}_{k,j}$ and $\delta_{i,j} =$

$(\widetilde{H}_j - \widetilde{H}_i)_+ = \langle {}^{\widetilde{H}_j - \widetilde{H}_i \; if \; \widetilde{H}_i < \widetilde{H}_j}_{0 \quad otherwise}, \widetilde{\delta}_{k,j} = \sum_{\substack{j=1 \\ j \notin q_i}}^{N-q_i} \frac{(\widetilde{H}_k - \widetilde{H}_j)_+}{N}$.[20] Consequently,

$$\begin{aligned} DMIM &= \sum_{i=1}^{L} \phi_i \sum_{i=1}^{L} \phi_i G_i + \widetilde{H}[\sum_{i=1}^{L}[\phi_i^2 \frac{H_i}{\widetilde{H}} \frac{\sum_{k=1}^{q_i} \overline{\delta}_{k,j}}{H_i q_i}] + \sum_{i=1}^{L} \sum_{k=1}^{q_i} \frac{\widetilde{\delta}_{k,j}}{\widetilde{H}N}] \\ &= \sum_{i=1}^{L} \phi_i [\phi_i G_i + \sum_{\substack{j=1 \\ j \neq i}}^{L} \phi_j G_j] + \sum_{i=1}^{L} \phi_i^2 H_i I_i + \widetilde{H} \sum_{i=1}^{L} \sum_{k=1}^{q_i} \frac{\widetilde{\delta}_{k,j}}{\widetilde{H}N} \\ &= \sum_{i=1}^{L} \phi_i^2 (G_i + H_i I_i) + \widetilde{H} \sum_{i=1}^{L} \sum_{k=1}^{q_i} \frac{\widetilde{\delta}_{k,j}}{\widetilde{H}N} + \sum_{i=1}^{L} \phi_i \sum_{\substack{j=1 \\ j \neq i}}^{L} \phi_j G_j] \end{aligned}$$

Recalling that the membership of each group is distinct, the element of the above expression which corresponds to the Gini coefficient of individual health status can be rewritten as:

$$\begin{aligned} \sum_{i=1}^{L} \sum_{k=1}^{q_i} \frac{\widetilde{\delta}_{k,j}}{\widetilde{H}N} &= \sum_{i=1}^{L} \sum_{k=1}^{q_i} \frac{\sum_{\substack{j=1 \\ j \notin q_i}}^{N-q_i} \frac{(\widetilde{H}_k - \widetilde{H}_j)_+}{N}}{\widetilde{H}N} = \sum_{i=1}^{L} \sum_{k=1}^{q_i} \frac{\sum_{\substack{j=1 \\ j \neq i}}^{L} \phi_j (\overline{H_k} - \overline{H_j})_+}{\widetilde{H}N} \\ &= \sum_{i=1}^{L} \sum_{k=1}^{q_i} \frac{\sum_{\substack{j=1 \\ j \neq i}}^{L} \phi_j (\overline{H_k} - \overline{H_j})_+ + \phi_k (\overline{H_k} - \overline{H_k})}{\widetilde{H}N} \\ &= \sum_{i=1}^{L} \sum_{k=1}^{q_i} \frac{\sum_{j=1}^{L} \phi_j (\overline{H_k} - \overline{H_j})_+}{\widetilde{H}N} = \sum_{i=1}^{L} \sum_{k=1}^{q_i} \frac{\frac{\sum_{p=1}^{N} (\overline{H_k} - H^p)_+}{N}}{\widetilde{H}N} \\ &= \sum_{i=1}^{N} \frac{\overline{\delta^i}}{\widetilde{H}N} = \overline{I}. \end{aligned}$$

Now, defining $H^p = \overline{H_i}$ when $p \in q_i$,i.e supposing that every individual takes the average health experience of the sub-group members as his health status, then we can re-express the DMIM as

$$\begin{aligned} DMIM &= \sum_{i=1}^{L} \phi_i^2 DMIM_i + \widetilde{H}\overline{I} + \widetilde{G} - \sum_{i=1}^{L} \phi_i^2 G_i \\ &= \sum_{i=1}^{L} \phi_i^2 DMIM_i + \overline{DMIM} - \sum_{i=1}^{L} \phi_i^2 G_i \end{aligned}$$

[20] Also see Araar (2006) for further detail.

■

Conceptually, the $DMIM$ approach is similar to the method developed by Cruces, Makdissi & Wodon (2006) to measure poverty under risk aversion using panel data. Their method also has two steps, where in the first step the risk-adjusted over-time income measures at the individual level are computed, then in the second step, these computed incomes are used to estimate the risk-adjusted poverty measures. The effect from averaging income fluctuations which mitigate the impact of shocks interacts with the effect of disutility from income fluctuations in the presence of risk aversion, with the latter eventually dominating as risk aversion increases. In contrast, $DMIM$ records the individual's average health experience over time and adjusts for the fluctuations in his or her health status. But when a more health inequality averse preference is adopted, the negative effect on population health is not coming from the fluctuation in health over time but rather from the disparity of individuals' health experiences.

There is another noteworthy difference between the Cruces, Makdissi and Woden approach and the $DMIM$. In the former's method, any case with reported zero income at any point in time cannot be included in the estimation, which is somewhat at odds with the notion that the length of poverty matters when individuals are risk averse. Whereas in the calculation of $DMIM$ any individual with zero health (death) at any point in time will be kept in the calculation, and this individual's health level is equal to 0 thereafter.

3.5 Alternative Approaches to the Dynamic Measure of Population Health

As described above, our purpose of this chapter is to propose a reasonable candidate approach for the measurement of population health over time, instead of defining the only ideal approach. Although there are no existing alternative dynamic measures to which the $DMIM$ may be compared, there are a number of potential alternatives, and it is very helpful to compare the $DMIM$ with these other possible dynamic measures of population health.

3.5.1 Alternative MIM-Based Dynamic Measure of Population Health

Trend version of MIM ($TMIM$)

The trend-based version of MIM is also a two step MIM. It is constructed by computing the MIM value for each time period and then aggregating these measures (again using the MIM approach).

$$TMIM(X) = [\alpha \widehat{G}^{\rho} + \beta(\widehat{H}\widehat{I})^{\rho}]^{1/\rho}$$

where $\overline{G_t} = \frac{\sum_{i=1}^{n} G_{i,t}}{n}$, $\overline{H_t} = 1 - \frac{\sum_{i=1}^{n} G_{i,t}}{n}$, $I_t = I(h_{1,t}, ..., h_{n,t})$, $\widetilde{G_t} = \overline{G_t} + \overline{H_t} I_t$, and $\widehat{H} = 1 - \widehat{G}$, and $\widehat{G} = \frac{\sum_{t=1}^{T} \widetilde{G_t}}{T}$, $\widehat{I} = I(\widetilde{H_1}, ..., \widetilde{H_T})$, $\rho \leq 1$.

Whereas $TMIM$ is designed to capture the dynamics of population health, irrespective of the dynamics of health profile of each specific individual, it ignores the relationship of health distributions of different periods. It satisfies the Weak Pareto axiom, Replication

Invariance, and Transfer Sensitivity, but it does not satisfy the Weak Anonymity axiom, nor does it exhibit the property of decomposability.

Average MIM ($AMIM$)

As a variant of the $TMIM$, the $AMIM$ is obtained if we calculate out the MIM value of the population health for each wave and take a simple mathematical average. Similarly to $TMIM$, $AMIM$ satisfies the Weak Pareto axiom, Replication Invariance, and Transfer Sensitivity axiom, but it fails to satisfy the Weak Anonymity axiom and does not exhibit the property of decomposability.

3.5.2 Other Potential Alternative Approaches

It is interesting to compare the approach taken here to Foster's (2007) recently proposed approach to the measurement of chronic poverty which - at a conceptual level at least - is obviously a somewhat similar problem. Adapting Foster's proposed measure to the health context, and defining $g_{i,t}$ as the gap between individual i's actual level of health in period t and perfect health, then a 'chronic health poverty' measure can be constructed as an average of the weighted exponential sumer of individual health gaps, i.e.,

$$MCM = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{Tg_{i,t}} g_{i,t}^{\alpha}$$

where $\alpha > 2$. The MCM measure satisfies monotonicity, replication invariance, transfer sensitivity, and decomposability, but once again does not satisfy the axiom of Weak Anonymity as described above. In effect, therefore, it is not sensitive to changes in the distribution of health experience of individuals over time.

Similarly, comparing DMIM to more traditional measures, we note that - as discussed in Chapter 1 - summary measures of population health such as the DALY and HALY merely sum up the years of average health experience of individuals. In effect, what a conventional measure (CM) does is to measure population health as an average of individual DALY or HALY scores, i.e.,

$$CM = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} h_{i,t}.$$

The CM captures the mean of individuals' average health experience over their lifetime. It satisfies monotonicity, replication invariance, and decomposability, but doesn't satisfy either transfer sensitivity or the principle of weak anonymity, as expressed above. Therefore, it does not give specific consideration to the distribution of either population health or individual health over time, and therefore does not capture the problem of health inequality in the measurement of population health over time.

3.6 Application: The China Health and Nutrition Survey

The theoretical analysis above has provided a careful analysis of the proposed $DMIM$ for measuring the health of a population over time. Obviously, however, it is of interest to examine the insights that can be gleaned from actually applying this approach to the analysis of actual data, and to compare the results obtained using this approach to some of the potential alternatives. Therefore, for illustrative purposes, in this section we apply the $DMIM$ to the China Health and Nutrition Survey (CHNS) longitudinal dataset[21], and

[21] The data is freely downloadable from http://www.cpc.unc.edu/china.

also compare the results generated with the $DMIM$ to what would be obtained using an average MIM or the $TMIM$, or CM and MCM.

The CHNS is conducted as part of an ongoing international collaborative project between the Carolina Population Center of the University of North Carolina at Chapel Hill and the National Institute of Nutrition and Food Safety of the Chinese Center for Disease Control and Prevention. It was designed to examine how the social and economic transformation of Chinese society is affecting the health and nutritional status of its population. An international multidisciplinary team of researchers conducts the survey using a multi-

stage, random cluster process to draw a sample of about 4400 households with a total of 19,000 individuals in nine provinces that vary substantially in geography, economic development, public resources, and health indicators. A select number of master files suitable for longitudinal analysis have been constructed by consolidating and standardizing data from multiple survey years.

We use the Physical Exam master file for our application. Although it includes demographic and physical exam data for surveyed individuals for all seven waves of the CHNS, for the purposes of the exercise conducted here only the most recent four waves (1997, 2000, 2004, 2006) are of interest, as there are missing values for the health status variable in the earlier waves. We transformed the master file into a balanced panel that contains 4532 individuals across four waves with observations for each wave. The health of each individual at each wave is measured by the self-reported health status on a scale from 1 to 4, where 1 represents excellent health and 4 indicates poor health. After normalization, we

transform the measure of self-reported health to a scale from 0 to 1, where 0 is the least healthy status and 1 is excellent health.

Using this data, we then compute the $DMIM$ value for the whole population. To illustrate the importance of the Weak Anonymity axiom, we also calculate the value of $TMIM$, $AMIM$, CM and MCM, and compare them to the $DMIM$ to check the difference. And next we apply the comparison of approaches to the rankings of health over subpopulations with respect to gender, urban/rural differences. Subsequently, we investigate the decomposability of the $DMIM$ measure by comparing the overall result for the population with the results obtained for subpopulations based on gender difference.

3.6.1 Comparing the DMIM, Trend MIM, Average MIM, CM and MCM

The first step in calculating the $DMIM$ is to compute the MIM value for each individual taking account of the distribution of individual's health experience across each of the four waves. At this stage, we require that $\rho = 1$. Recalling that $MIM_i = 0$ when the individual is in perfect health in each period, then we can construct a variable of individual health = $1 - MIM_i \in [0, 1]$. Figure 3.1 displays the cumulative distribution of individual health after this first stage calculation; individuals with a score of '1' are in perfect health in each period.

Subsequently, in the second stage, we calculate the $DMIM$ value for the whole population. The average individual life adjusted health at the second stage is 0.60710296, and the Gini coefficient of measure of health inequality of the whole population is 0.12378025.

Therefore it is straightforward to calculate the $DMIM$ value for different specifications of ρ, as reflected in figure 3.2.

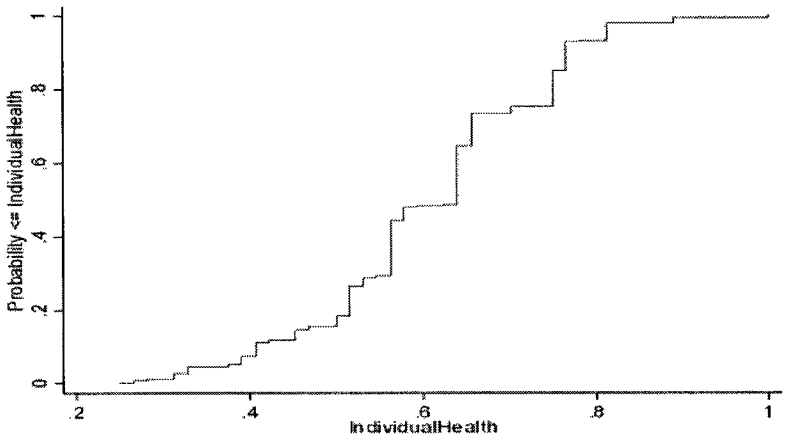

Fig. 3.1. CDF of individuals' health

	DMIM		DMIM
p=1	0.46804439	p=-1/10	0.00016217
p=1/2	0.8117022	p=-1/7	0.00127855
p=1/3	1.5384441	p=-1/5	0.0050162
p=1/4	2.992866	p=-1/3	0.01919158
p=1/5	5.8859674	p=-1/2	0.03637439
p=1/6	11.640009	p=-1	0.06308199
p=1/7	23.092751	p=-3	0.0749729
p=1/8	45.90601	p=-5	0.07514351
p=1/9	91.379003	p=-7	0.07514725
p=1/10	182.06746	p=-9	0.07514735

Fig. 3.2. DMIM Value of the Population for Different ρ

We observe that $DMIM$ is negatively correlated with ρ, and that the graph is discontinuous at $\rho = 0$. From the discussion in Chapter Two, we know that ρ is positively related to the substitution effect between the increase of average health and increase of health

inequality, and hence negatively related to the aversion towards health inequality. As ρ becomes smaller, the inequality aversion embedded in preferences increases. Consequently, the changes in the value of the $DMIM$ put increasing weight on the health experience of the less healthy members of the population.

For the purpose of better understanding the strength of the $DMIM$ approach as compared to a less truly dynamic approach to the measurement of population health over time, it is useful to compare the $DMIM$ with both the Trend MIM ($TMIM$) and with the average of the Annual MIM ($AMIM$).

To obtain $TMIM$, we first compute the population health for each wave without considering the sequential health experience of each individual, assuming as before that the ρ in the first stage is equal to 1. The MIM values for each of the four waves are reported in figure 3.3.

Year	Mean	Gini	MIM
1997	0.704325	0.124579	0.383419
2000	0.679667	0.14196	0.416818
2004	0.651092	0.16041	0.453349
2006	0.639508	0.159966	0.462792

Fig. 3.3. MIM Measure for Each Survey Wave

Then, in the second stage, the MIM is calculated for the distribution of the the yearly MIM values, for different possible values of ρ. The results of these calculations are reported in figure 3.4.

	TMIM		TMIM
p=1	0.44626	p=-1/10	0.00007367
p=1/2	0.617907	p=-1/7	0.00055814
p=1/3	1.037092	p=-1/5	0.00207902
p=1/4	1.882338	p=-1/3	0.00710221
p=1/5	3.54285	p=-1/2	0.01192032
p=1/6	6.798643	p=-1	0.01650528
p=1/7	13.19671	p=-3	0.01716519
p=1/8	25.80365	p=-5	0.01716556
p=1/9	50.70355	p=-7	0.01716556
p=1/10	99.97886	p=-9	0.01716556

Fig. 3.4. TMIM Value of the Population for Different ρ

Compared with $DMIM$, for every ρ, $TMIM$ has a lower value, which may possibly suggest that the population is healthier; however, because the qualitative rankings agree, it is not evident that these two approaches provide substantially different guidance to health policy makers. In contrast, because $TMIM$ violates the Weak Anonymity Axiom by disconnecting the health experiences in different periods of the same individual, it would appear to underestimate the actual experience of health inequality, because the actual fluctuation in the individual's health experiences is effectively submerged in the series of static measures of changes in the health of the population. Therefore, as greater weight is accorded to health inequality (via decreasing values of ρ), the $TMIM$ takes on values that deviate proportionally from $DMIM$ to a larger and larger extent, hence ultimately leading to overestimation of the population health.

To compute $AMIM$, we first calculate the MIM of the population health for each wave for different ρ, also ignoring the sequence of individual's health experience. Then,

taking the arithmetic mean, we have the $AMIM$ values for different ρ reported in figure 3.5.

	Average MIM		Average MIM
p=1	0.42909467	p=-1/10	0.00017249
p=1/2	0.788997943	p=-1/7	0.00136898
p=1/3	1.531241325	p=-1/5	0.005418235
p=1/4	3.016019225	p=-1/3	0.021148058
p=1/5	5.9766078	p=-1/2	0.041043205
p=1/6	11.8796425	p=-1	0.07547048
p=1/7	23.65454875	p=-3	0.09691774
p=1/8	47.1524245	p=-5	0.09769866
p=1/9	94.0616715	p=-7	0.097739865
p=1/10	187.734965	p=-9	0.097742433

Fig. 3.5. AMIM Value of the Population for Different ρ

Like the $TMIM$, the $AMIM$ also violates the Weak Anonymity Axiom, and is therefore relatively insensitive to variation in the lifetime health experience of individuals. As compared with $DMIM$, for relatively large values of ρ the $AMIM < DMIM$ but the gaps shrink as ρ gets smaller, and eventually become larger than $DMIM$, with the difference between these two measures increasing monotonically as ρ gets smaller. However, as before, it should again be observed that in this particular case the ordinal rankings are the same for both measures. By comparing health inequality and mean health for each wave (shown in figure 3.3) with the health inequality and mean health used in computing $DMIM$, we find that the health inequality of the population in each wave is larger than the health inequality captured in the distribution of the health of individuals after the first stage of the $DMIM$ calculation. The mean health of the population of each wave is also larger than the mean health of the distribution of the health of individuals' after the first stage of the $DMIM$. Therefore, by monotonicity, when $\rho = 1$ the value of the $AMIM$ will be

lower than that of the $DMIM$ because of higher average health. However, as ρ becomes smaller, $AMIM$ will eventually take on larger values than does the $DMIM$, because of the higher health inequality in every wave.

The relationship between $DMIM$, $TMIM$ and $AMIM$ is displayed by figure 3.6..

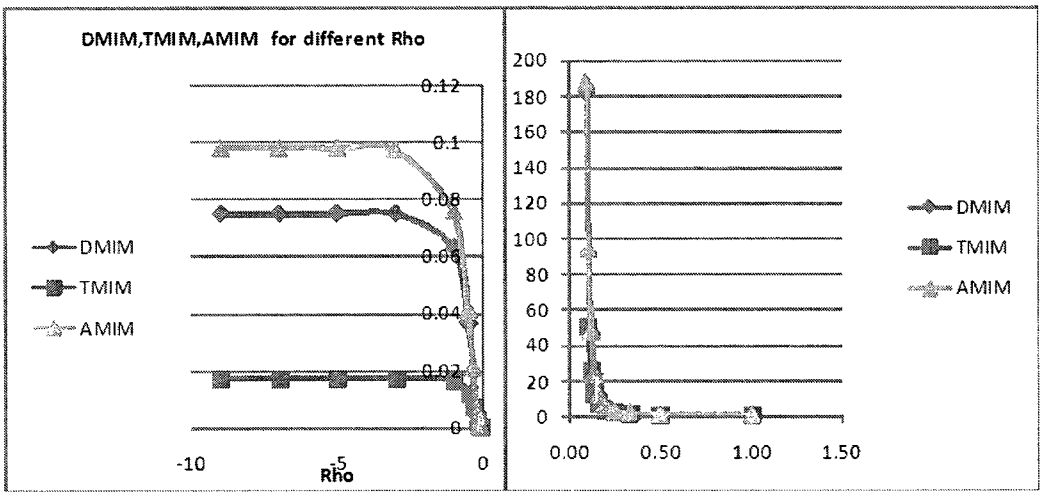

Fig. 3.6. The Comparison of DMIM, TMIM and AMIM for Different ρ

Meanwhile, we compute CM , -the value for the average health gap over time across individuals-, it is 0.3313521. It is much less than the MIM based measures, because it fails to account for the fluctuation of individual health over different periods, it is likely to overestimate the population health.

It is difficult to compare the MCM - the average exponential health gap of individuals over time - to the CM and other MIM based measures, although it is easy to calculate out its value 0.1457655. Because the parameters adopted in MCM changes the domain of the scale, it is not comparable with that generated by the application of ρ in MIM based measures. The way in which MCM differs from the other measures is revealed more

clearly in comparing the differences in the way the different measures rank the health of different populations.

3.6.2 Comparing the Health of Different Sub-Populations

In this section we show that the when $DMIM, AMIM$ and $TMIM$ are used to rank the health of different sub-populations that, depending on the actual value of ρ, they may generate quite different ordinal rankings. For the purposes of comparison, these rankings are also compared to those generated by CM and MCM. This exercise clearly illustrates the fact that these alternative approaches are likely to lead to quite different approaches to health policy making.

Figure 3.7 presents the $DMIM, TMIM$ and $AMIM$ calculations separately for males and females. The mean health of males across the four waves is 0.6258979 and the measure of health inequality of individuals is 0.11822954 , while the mean health of female is 0.59053938 and health inequality of individuals is 0.12664864. As males consistently dominate females both with respect to their mean health level and with respect to health inequality, it follows immediately that the $DMIM$ of males always dominates that of females, for all values of ρ. Importantly, as is reflected in figure 3.8, the pattern of higher mean health and lower health inequality for males as compared to females prevails for each of the four waves. Consequently, when we use $TMIM$ or $AMIM$ to compare the health of the male and female sub-populations, we generate the same ordinal rankings as when using $DMIM$: the male subpopulation is always found to be healthier than the female

subpopulation. Such conclusion is agreed by the ranking generated when using CM or MCM.

	DMIM		TMIM		AMIM	
	Male	Female	Male	Female	Male	Female
p=1	0.44810172	0.48425162	0.42620146	0.46289492	0.410018078	0.44500737
p=1/2	0.78086839	0.83424575	0.58911834	0.64133367	0.757481885	0.814062813
p=1/3	1.4830327	1.576669	0.98760168	1.076866	1.47272835	1.576756025
p=1/4	2.8881977	3.0625824	1.7912046	1.9550341	2.90346875	3.102483575
p=1/5	5.6839002	6.0174677	3.3697187	3.6802996	4.99424145	6.14411465
p=1/6	11.245447	11.892585	6.4642649	7.0632367	11.4471285	12.207485
p=1/7	22.317172	23.583177	12.544613	13.711492	22.799529	24.30005225
p=1/8	44.375014	46.864969	24.524059	26.81196	45.45733225	48.42823425
p=1/9	88.348185	93.26311	48.182191	52.687578	90.69465825	96.5897445
p=1/10	176.05545	185.78189	94.996041	103.89534	181.0379025	192.75355
p=0						
p=-1/10	0.00015724	0.00016484	0.00006985	0.00007662	0.00016672	0.000176655
p=-1/7	0.00124045	0.00129855	0.00052895	0.00058054	0.001323835	0.001401253
p=-1/5	0.0048705	0.00508918	0.00196915	0.00216289	0.005242963	0.00554199
p=-1/3	0.01866676	0.01942321	0.00671878	0.00739191	0.020494593	0.021595505
p=-1/2	0.03545209	0.03670857	0.0112634	0.01241179	0.039846753	0.041828788
p=-1	0.06177931	0.06323979	0.01556887	0.01719634	0.07361676	0.076522508
p=-3	0.07380969	0.07463969	0.01618304	0.01788718	0.095283753	0.097472533
p=-5	0.07399514	0.07478797	0.01618338	0.01788756	0.096140815	0.098172678
p=-7	0.0739995	0.07479094	0.01618338	0.01788756	0.096189745	0.098206805
p=-9	0.07399962	0.07479101	0.01618338	0.01788756	0.09619304	0.098208773

Fig. 3.7. Comparison of the Health of Male and Female of the Population Using DMIM, TMIM and AMIM

	Mean Health		Health Inequality	
	Male	Female	Male	Female
1997	0.71773434	0.69250726	0.11783734	0.13003251
2000	0.69736222	0.66407223	0.13682488	0.14538879
2004	0.66957136	0.63480697	0.15469551	0.16410384
2006	0.66003297	0.62141968	0.15332737	0.16424367

Fig. 3.8. Mean Health and Health Inequality for of Male and Female for Each Wave

In contrast, if the population is divided into Rural and Urban sub-populations, differences in the rankings emerge when we use $DMIM$, $TMIM$ and $AMIM$. The results are collected and shown in figure 3.9.

	DMIM		TMIM		AMIM	
	Urban	Rural	Urban	Rural	Urban	Rural
p=1	0.47718663	0.46364517	0.45399261	0.44263315	0.440422343	0.423713858
p=1/2	0.81916273	0.80729435	0.60861016	0.62170153	0.804947373	0.780954128
p=1/3	1.5457859	1.532691	0.99807018	1.0535793	1.558555575	1.51700795
p=1/4	3.0001362	2.9843606	1.7854941	1.9235632	3.06611585	2.989387175
p=1/5	5.8918022	5.8724863	3.3287706	3.6343638	6.071422125	5.925523225
p=1/6	11.640268	11.61768	6.3455613	6.9928753	12.0621605	11.780357
p=1/7	23.077183	23.054631	12.257412	13.600229	24.009493	23.4600715
p=1/8	45.850997	45.839483	23.878193	26.632152	47.8472545	46.7695815
p=1/9	91.232151	91.260942	46.783256	52.392601	95.42800625	93.305457
p=1/10	181.71525	181.85518	92.031753	103.40647	190.4304375	186.2376425
p=0						
p=-1/10	0.00016089	0.00016235	0.00006494	0.00007752	0.000174445	0.000171315
p=-1/7	0.00126692	0.0012806	0.00048759	0.00058943	0.001383605	0.00135998
p=-1/5	0.0049623	0.00502746	0.00179545	0.00220572	0.005471473	0.00538438
p=-1/3	0.01891395	0.01926265	0.0059882	0.0076087	0.02131438	0.021031838
p=-1/2	0.03569122	0.03657116	0.00982016	0.01289424	0.041269375	0.040854943
p=-1	0.06126935	0.06367734	0.01316464	0.01811064	0.07542712	0.07530429
p=-3	0.07205485	0.07600892	0.01357013	0.01891874	0.095922955	0.097092175
p=-5	0.07218803	0.0761967	0.01357026	0.0189193	0.09659406	0.097920798
p=-7	0.07219056	0.07620106	0.01357026	0.0189193	0.096626035	0.097966428
p=-9	0.07219062	0.07620118	0.01357026	0.0189193	0.096627838	0.097969393

Fig. 3.9. Comparison of the Health of Urban and Rural People Using DMIM, TMIM and AMIM

The $DMIM$ of urban people is larger than that of rural at the beginning when $\rho = 1$, but the ranking switches for values of ρ less than1/9. Observe that the mean health of urban people is 0.595004, and the health inequality measure for individuals is 0.12132796, whereas the mean health of rural people is 0.61255602, the health inequality measure for individuals is 0.12439872. As the value of ρ decreases, health inequality is weighted more heavily in the calculation of the $DMIM$, and consequently urban people are found to be healthier than rural ones because there is less inequality in the distribution of health amongst urban residents.

Considering figure 3.10, which reports the mean health and the level of health inequality for urban and rural people for each of the four waves of the survey, we observe

that the mean health level of urban people is consistently lower than that of rural people for all four waves, but that the there is not a consistent ranking of health inequality across each wave: health inequality was less for urban residents than for rural residents in 2000 and 2004, whereas it was higher for 1997 and 2006. It is therefore not surprising to find that, depending on the value of ρ, the ordinal rankings of the health of the urban and rural populations differ when using $TMIM$ and $AMIM$ rather than $DMIM$. The $TMIM$ measure ranks the rural population as healthier than the urban one for ρ values less than 1/2, whereas it is not until ρ equals -3 that the $AMIM$ ranks the health of the two populations in the same way as the $DMIM$.Figure 3.9 above illustrates our earlier discussion, which pointed out that - because it neglects the inequality in the distribution of health at the level of individuals - the $TMIM$ will tend to overestimate the health of a population.

	Mean Health		Health Inequality	
	Urban	Rural	Urban	Rural
1997	0.682706	0.714069	0.130114	0.12127
2000	0.665661	0.68598	0.141039	0.141886
2004	0.642578	0.65493	0.158319	0.161203
2006	0.633878	0.642045	0.161018	0.159453

Fig. 3.10. Mean Health and Health Inequality of Urban and Rural People for Each Wave

The ranking over urban and rural population health does not change when using CM, the value of CM for urban population is 0.3437944, whereas that for rural population is 0.3257442, therefore, the conclusion reached through comparison of CM is that the rural people are always healthier than urban people. MCM ranks the urban and rural populations in the same way as CM, regardless of which α is adopted as the power of the health gap. Because the tradeoff of average health and health inequality is not specifically

incorporated into either the CM or MCM, it is not surprising to see the invariant ranking of the two sub-populations.

3.6.3 Decomposing DMIM for Gender Groups

When $\rho = 1$, we can study the decomposability of $DMIM$. From figure 3.7, we know that the $DMIM$ for males is 0.44810172, whereas the $DMIM$ for females is 0.48425162. The share of males in the population is 46.84% while that of females is 53.16%. Recall that the $DMIM$ of the whole population is 0.468. By the decomposability equation, we have $DMIM = \sum_{i=1}^{L} \phi_i^2 DMIM_i + \overline{DMIM} - \sum_{i=1}^{L} \phi_i^2 G_i$. We can substitute these values into the right hand side of the decomposability equation and compare this to the overall $DMIM$. This gives us: $0.4684^2 * 0.4481 + 0.5316^2 * 0.48425 + 0.429997 - 0.4684^2 * 0.3741021 - 0.5316^2 * 0.40946062 = 0.46737$, which is very close to the value of 0.468 for the $DMIM$ (and the difference may be attributed to rounding error). This demonstrates that, in interpreting the value of the $DMIM$, it is indeed appropriate to view this measure as decomposable into a between group component and within group component. In this particular case, the female group contributes more to the $DMIM$ of the population, while the between group $DMIM$ is less than that of either male or female group. The practical policy relevance of this finding is that it enables better targetting of health policy initiatives. In particular, in this context it is clear that priority should be given to implementing policies to improve the health of females.

3.7 Conclusions

The analysis undertaken in this chapter proposes a framework for thinking about how to measure population health over time, and proposes a new summary measure of population health, the $DMIM$. Like its static counterpart, the MIM measure, the $DMIM$ is sensitive to differences in the distribution of health across individuals, and provides a means to explicitly take account of the tradeoff between improvements in the average level of health, and improvements in the distribution of health across individuals. In developing the $DMIM$ measure, we draw upon the existing literatures on income inequality and poverty to inform the selection of a set of axioms which it is argued should be satisfied by a satisfactory measure of dynamic population health. This exercise allows us to identify a set of desirable axioms that should be satisfied by a dynamic measure of population health - which include a weak version of Pareto sensitivity, a measure of replication invariance, weak anonymity, and transfer sensitivity - and we construct a measure - the $DMIM$ - which satisfies these axioms. We also show that alternative potential approaches $(AMIM, TMIM, CM, MCM)$ to generating an inter-termporal measure of population health do not satisfy these axioms. We illustrate the use of $DMIM$ by applying it to the CHNS longitudinal health data, and compare the $DMIM$ measure with these alternative approaches. This comparison provides a convincing illustration of the strength of the $DMIM$ approach.

References

Anand, S. and Hanson, K. (1997) Disability-adjusted life years: a critical review, *Journal of Health Economics*, **16**: 685-702.

Araar, A. (2006) On the decomposition of the Gini coefficient: an exact approach, with an illustration using Cameroonian data, Centre Interuniversitaire sur le Risque, les Politiques Economiques et l'Emploi (CIRPEE) working paper 06-02.

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

Atkinson, A. B., and Bourguignon, F. (1982) The comparison of Multi-dimensioned distributions of economic status, *The Review of Economic Studies*, **49(2)**: 183-201.

Blackorby, C., and Donaldson, D. (1978) Measures of relative inequality and their meaning in terms of social welfare, *Journal of Economic Theory,* **18**: 59-80.

Bleichrodt, H., and van Doorslaer, E. (2006) A welfare economic foundation for health inequality measurement, *Journal of Health Economics*, **25**: 945-957.

Bleichrodt, H., and Filko, M. (2008) New tests of QALYs when health varies over time, *Journal of Health Economics*, **27**: 1237-1249.

Bommier, A., and Stecklov, G. (2002) Defining health inequality: why Rawls succeeds where social welfare theory fails, *Journal of Health Economics,* **21**: 497–513.

Bourguignon, F., and Chakravarty, S. (2003) The measurement of multidimensional poverty, *Journal of Economic Inequality*, **1(1)**: 25-49.

Brazier, J., Deverill, M. and Green, C. (1999) A review of the use of health status measures in economic evaluation, *Journal of Health Services Research and Policy*, **4(3)**: 174–184.

Brazier, J., Roberts, J., and Deverill, M. (2002) The estimation of a preference-based measure of health from the SF-36, *Journal of Health Economics,* **21**: 271-292.

Brown, M. (1994) Using Gini-style indices to evaluate the spatial patterns of health practitioners: theoretical considerations and an application based on Alberta data, *Social Science and Medecine,* **38(9)**: 243-1256.

Calvo, C., and Dercon, S. (2007) Chronic poverty and all that: the measurement of poverty over time, Chronic Poverty Research Centre (CPRC) working paper 89.

Cambois, E., J. M. Robine, and N. Brouard (1998) Life expectancies applied to specific statuses: A history of the indicators and the methods of calculation, *Population,* **53(3)**: 447-476.

Chiu, W. H. (2007) Intersecting Lorenz curves, the degree of downside inequality averstion, and tax reforms, *Social choice and Welfare,* **28**: 375-399.

Clarkea, P. M., Gerdthamb, U. G., Johannessonc, M., Bingeforsd, K., and Smithe, L. (2002) On the measurement of relative and absolute income-related health inequality, *Social Science and Medicine,* **55**: 1923–1928.

Contoyannis, P., and Forster, M. (1999) The distribution of health and income: a theoretical framework, *Journal of Health Economics*, **18**: 605–622.

Cruces, G., Makdissi, P., and Wodon, Q. (2006) Poverty measurement under risk aversion using panel data, University of Sherbrooke, Working paper: 03-06

Dachs, N. (2002) An equity motivated indicator of population health, In *Summary Measures of Population Health: Concepts, Ethics, Measurements and Applications*, by Murray, C. J. L., Salomon, J. A., Mathers, C. D., and Lopez, A. D., Eds., Geneva: World Health Organization, 161-169.

Davidson, R., and Duclos, J. (2000) Statistical inference for stochastic dominance and for the measurement of poverty and inequality, *Econometrica,* **68(6)**: 1435-1464.

Davies, J., and Hoy, M. (1995) Making Inequality Comparisons when Lorenz curves intersect, *American Economic Review,* **85**: 980-986.

Dolan, P. (2000) The measurement of health-related quality of life for use in resource allocation decision in health care, In *Handbook of Health Economics* by Culyer, A. J., and Newhouse, J. P., Eds., Amsterdam: North-Holland, **1**: 1723-1760.

Dolan, P., and Gudex, C. (1995) Time preference, duration and health state valuations, *Health Economics,* **4**: 289-299.

Duclos, J., Sahn, D., and Younger, S. D. (2006) Robust multi-dimensional poverty comparisons, *Economic Journal,* **116**: 943-968.

Dutta, B. (2002) Inequality, poverty and welfare, In *Handbook of Social Choice and Welfare,* by Arrow, K. J., Sen, A. K., and Suzumura, K., Eds., Amsterdam: North-Holland, **1**: 597-633.

Feeny, D., Furlong, W., Torrance, G. W., *et al* (2002) Multi-attribute and single-attribute utility functions for the health utilities index mark 3 system, *Medical Care*, **40(2)**: 113-128.

Fishburn, P. C (1976) Continua of stochastic dominance relations for bounded probability distributions, *Journal of Mathematical Economics,* **3**: 295-311.

Foster, J. E., Greer, J., and Thorbecker, E. (1984) A class of decomposable poverty measures, *Econometrica,* **52:** 761-766.

Foster, J. E, (1985) Inequality Measurement, In *Fair Allocation* by Young, H. P., Eds., Providence, Rhode Island: American Mathematical Society, 31-68.

Foster, J. E, and Shorrocks, A. F. (1988) Inequality and poverty ordering, *European Economic Review,* **32**: 654-662.

Foster, J. E. (2005) Poverty Indices, In *Poverty, Inequality and Development: Essays in Honor of Erik Thorbecke,* by de Janvry, A., and Kanbur, R., Eds., Amsterdam: Kluwer Academic Publishers, 41-65.

Foster, J. E. (2007) A Class of Chronic Poverty Measures, Vanderbilt University, Department of Economics, working paper No.07-W01.

Froberg, D. G., and Kane, R. L. (1989) Methodology for measuring health-state preferences part – II: scaling methods, *Journal of Clinical Epidemiology,* **42(5)**: 459–471.

Gajdos, T. (2004) Single crossing Lorenz curves and inequality comparisons, *Mathematical Social Sciences,* **47**: 21-36.

Gakidou, E. E., Murray, C. J. L., and Frenk, J. (2000) Defining and measuring health inequality: An approach based on the distribution of health expectancy, *Bulletin of the World Health Organizaition,* **78(1)**: 42-54.

Hammitt, J. K. (1993) Discounting health increments, *Journal of Health Economics,* **12**: 117-120.

Health Canada (2001) *The population health template: Key elements and action that define a population health approach*, Ottawa, Canada.

Jalan, J., and Ravallion, R. (1998) Determinants of transient and chronic poverty, evidence from rural China, The World Bank, Policy Research working paper No.1936.

Jones, A. M., and Nicolas, A. L. (2004) Measurement and explanation of socioeconomic inequality in health with longitudinal data, *Health Economics,* **13**: 1015-1030.

Kakwani, N. C. (1977) Applications of Lorenz curves in Economics Analysis, *Econometrica,* **45(3)**: 719-728.

Kakwani, N. C. (1980) *Income inequality and Poverty: Methods of estimation and policy applications,* New York: Oxford University Press.

Kakwani, N. C., Wagstaff, A., and van Doorslaer, E. (1997) Socioeconomic inequalities in health: Measurement, computation and statistical inference, *Journal of Econometrics,* **77**: 87-103.

Kolm, S. (1969) The optimal production of social justice, In *Public Economics,* by Margolis, J., and Guitton, H., Eds, Macmillan: London, 145-200.

Koolman, X., and van Doorslaer, E. (2004) On the interpretation of a concentration index of inequality, *Health Economics,* **13**: 649-656.

Kunst, A. E., Geurts, J. J., and van den Berg, J. (1995) International variation in socioeconomic inequalities in self reported health, *Journal of Epidemiology and Community Health,* **49**: 117 – 123.

Lahelma, E., and Valkonen, T. (1990) Health and social inequalities in Finland and elsewhere, *Social Science and Medecine,* **31**: 257-265.

Lambert, P. (1989) *The distribution and redistribution of Income: A mathematical analysis*, Massachusetts: Basil Blackwell Inc.

Leclerc, A., Lert, F., and Fabien, C. (1990) Differential mortality: some comparisons between England and Wales, Finland and France, based on inequality measures, *International Journal of Epidemiology,* **9**: 1001-1010.

Le Grand, J., (1989) An international comparison of distribution of ages-at-death, In *Health Inequality in European Countries* by Fox, J., Eds, Aldershot, UK: Gower Publishing Company, 75-91.

Machenbach, J. P., and Kunst, A. E. (1997) Measuring the magnitude of socio-economic inequalities in health: an overview of available measures illustrated with two examples from Europe, *Social Science and Medicine,* **44(6)**: 757-771.

Molla, M., Madans, J., Wagener, D., and Crimmins, E. (2003) *Summary measures of population health: report of findings on methodologic and data issues,* Hyattsville, Maryland: National Center for Health Statistics.

Moyes, P., (1999) Stochastic dominance and the Lorenz curve, In *Handbook on income inequality measurement* by Silber, J., Ed., Boston: Kluwer Academic Publishers, 199-226.

Muliere, P., and Marco, S. (1989) A Note on Stochastic Dominance and Inequality Measures, *Journal of Economic Theory,* **49**: 314-323.

Murray, C. J. L., Gakidou, E. E., and Frenk, J. (1999) Health inequalities and social group differences: what should we measure?, *Bulletin of the World Health Organization,* **77**: 7.

Murray, C. J. L., and Acharya, A. (2002) Age weights and discounting in health gap reconsidered, In *Summary Measures of Population Health: Concepts, Ethics, Measurements and Applications*, by Murray, C. J. L., Salomon, J. A., Mathers, C. D., and Lopez, A. D., Eds., Geneva: World Health Organization, 677-684.

Pamuk, E. (1988) Social-class inequality in infant mortality in England and Wales from 1921 to 1980, *European Journal of Population,* **4**: 1-21.

Pearcy, J. N., and Keppel, K. G. (2002) A summary measure of health disparity, *Public Health Reports,* **117**: 273-280.

Pradhan, M., Sahn, D. E., and Younger, S. D. (2003) Decomposing world health inequality, *Journal of Health Economics,* **22**: 271-293.

Redelmeier, D. A., and Heller, D. N. (1993) Time preference in medical decision-making and cost-effectiveness analysis, *Medical Decision Making,* **13**: 212-217.

Regidor, E. (2004) Measures of health inequalities: part 2, *Journal of Epidemiology and Community Health,* **58**: 900-903.

Richardson, J., Hall, J., and Salkeld, G. (1996) The measurement of utility in multiphase health states, *International Journal of Technology Assessment in Health Care*, **12**: 151-162.

Rodgers, J. R., and Rodgers, J. L. (1993) Chronic poverty in the United States, *Journal of Human Resources,* **18**: 25-54.

Sen, A. K. (1970) On the measurement of inequality, *Journal of Economic Theory,* **2**: 244-263.

Sen, A. K. (1973) *On economic inequality,* Oxford: Clarendon Press.

Sen, A. K. (1976) Poverty: an ordinal approach to measurement, *Econometica,* **44**: 219-231.

Shorrocks, A. F. (1983) Ranking income distribution, *Economica,* **50**: 3-17.

Shorrocks, A. F, and Foster, J. E. (1987) Transfer Sensitive Inequality Meaures, *The Review of Economic Studies,* **54**: 485-497.

Spencer, A., Robinson, A. (2007) Test of utility independence when health varies over time, *Journal of Health Economics,* **26**: 1003-1013.

Stark, O., and Yitzhaki, S. (1988) Merging populations, stochastic dominance and Lorenz curves, *Journal of Population Economics,* **1**: 157-161.

Torrance, G. W., Boyle, M. H., and Horwood, S. P. (1982) Application of multi-attribute utility theory to measure social preferences for health states, *Operations Research,* **30(6)**: 1043-1069.

Torrance, G. W. (1986) Measurement of health state utilities for economic appraisal: A review, *Journal of Health Economics,* **5**: 1-30.

van Doorslaer, E., Wafstaff, A., Bleichrodt, H., *et al* (1997) Income-related inequalities in health: some international comparisons, *Journal of Health Economics,* **16**: 93-112.

van Doorslaer, E. and Jones, A. M. (2003) Inequalities in self-reported health: validation of a new approach to measurement, *Journal of Health Economics,* **22**: 61–87.

Wagstaff, A., van Doorslaer, E., and Paci, P. (1989) Equity in the finance and delivery of health care: some tentative cross-country comparisons, *Oxford Review of Economic Policy,* **5**: 89-112.

Wagstaff, A., Paci, P., and van Doorslaer, E. (1991) On the measurement of inequalities in health, *Social Science and Medicine,* **33**: 545-557.

Wagstaff, A., van Doorslaer, E. (2000) Equity in health care finance and delivery, In *Handbook of Health Economics* by Culyer, A. J., and Newhouse, J. P., Eds., Amsterdam: North-Holland, **1**: 1803-1862.

Wagstaff, A. (2002) Inequality aversion, health inequalities and health achievement, *Journal of Health Economics,* **21**: 627–641.

Wagstaff, A. and Watanabe, N. (2003) What difference does the choice of SES make in health inequality measurement?, *Health Economics,* **12**: 885-890.

Wagstaff, A., and van Doorslaer, E. (2004) Overall vs. Socioeconomic Health Inequality: A Measurement Framework and Two Empirical Illustrations, *Health Economics,* **13**: 297-301.

Weymark, J. A. (1999) Comment on Blackorby, Bossert and Donaldson, In *Handbook of Income Inequality Measurement* by Silber, J., Eds., Boston: Kluwer Academic Publishers, 157-161.

Wildman, J. (2003) Modelling health, income and income inequality: the impact of income inequality on health and health inequality, *Journal of Health Economics*, **22**: 521–538.

Wildman, J. (2003) Income related inequalities in mental health in Great Britain: analysing the causes of health inequality over time, *Journal of Health Economics,* **22**: 295–312.

Wolfson, M., and Rowe, G. (2001) On measuring inequalities in health, *Bulletin of the World Health Organization*, **79(6)**: 553-560.

Yaari, M. E. (1987) The dual theory of choice under Risk, *Econometrica,* **55(1)**: 95-115.

Yfantopoulos, J. (2001) Quality of life and QALYs in the measurement of health, *Archives of Hellenic Medicine*, **18(2)**: 114-130.

Zheng, B. (1997) Aggregate poverty measures, *Journal of Economic Surveys,* **11(2)**: 123-162.

Zoli, C. (1999) Intersecting generalized Lorenz curves and the Gini index, *Social choice and Welfare,* **16**: 183-196.