An Ordinal Measure of Population Health

Héctor Pifarré i Arolas^{1,2} · Christian Dudel¹

Accepted: 8 October 2018 / Published online: 12 October 2018
© The Author(s) 2018

Abstract

We propose a population-level health index that addresses two of the main concerns of existing health measures: (1) the biases in self-assessed health measures, and (2) the difficulties associated with making comparisons across populations afflicted by a variety of conditions in a context in which multi-morbidity is high. Starting from a set of general axioms, we derive a partial order index that ranks health across populations based on the relative prevalence of various conditions within these populations, and that can be readily applied to many existing health surveys. We illustrate the use of our health measure by applying it to data from the National Health Interview Survey in order to examine health differences across racial groups in the U.S.

Keywords Health disparities · Partial ordering · Comparability · NHIS

1 Introduction

Comparisons of health across populations are often made in the academic literature, and are frequently cited in political discussions. From cross-country evaluations (König et al. 2009), to studies on the socioeconomic gradient of health (Marmot 2005), to research on changes in health over time (Anderson et al. 2000), relative health assessments are among the most studied and hotly debated topics across disciplines. However, despite the importance of this issue, there is still an active debate about how to best operationalize population-level health comparisons (for a review of the literature, see Murray 2002 and Etches et al. 2006).

In this work we focus on the measurement of morbidity and introduce an approach specifically designed to tackle two of the main challenges faced by health indicators, as

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11205-018-2010-4>) contains supplementary material, which is available to authorized users.

✉ Héctor Pifarré i Arolas
pifarreirolas@demogr.mpg.de

Christian Dudel
dudel@demogr.mpg.de

¹ Max Planck Institute for Demographic Research, Konrad-Zuse-Strasse 1, 18057 Rostock, Germany

² Center for Research in Health and Economics, Universitat Pompeu Fabra, Barcelona, Spain

discussed by Burgard and Chen (2014). Namely, differences in disease distributions across groups and problems with self-assessed health measures. Health comparisons are often made across populations that are not necessarily afflicted by the same types of diseases. This issue of comparability is perhaps most obvious in cross-country comparisons among countries in different stages of the epidemiological transition and in health comparisons over time (Burgard and Chen 2014), but it is also present in subnational health assessments (Rivera et al. 2002; Weimann et al. 2016). A further potential problem that can arise when using existing measures is that some of the most frequently collected measures of health rely on self-assessments of health, despite a growing body of evidence that suggests that there is systematic bias in how respondents rate their health. While self-assessed health measures have been validated by their association with objective health indicators such as mortality (Idler and Benyamini 1997, 1999; Van Doorslaer and Gerdtham 2003; Jürges 2008), or health utilization (Humphries and Van Doorslaer 2000), more recent research has found evidence that certain socioeconomic groups suffer from reporting bias (Dowd and Zajacova 2010; Van Doorslaer et al. 2002; Etilé and Milcent 2006). Thus, the reliance on self-rated health measures when making group comparisons has been called into question.

Our strategy for addressing these challenges is to develop an analytical framework that is based strictly on relative prevalences. Comparisons based on disease prevalences have a long tradition in the public health literature, and comparisons that use this approach remain common (Braveman et al. 2010). The appeal of focusing on relative disease prevalences lies in the assumption that these indicators are more objective than self-rated health measures, and are more directly comparable in cross-population studies. Furthermore, as was noted in Burgard and Chen (2014), measures based on disease prevalences are more closely related to the origins of health disparities, and can therefore provide more direct guidance for public health interventions than other methods that aggregate health states in complex ways. We recognize that despite these advantages prevalence-based analyses are not without problems. As noted by Burgard and Chen (2014), diagnosis and reporting bias can lead to an over- or an underestimation of the relative disease burden across socioeconomic groups (Johnston and Propper 2009). While we acknowledge the importance of such issues, addressing such data collection concerns is beyond the scope of this work.

Setting aside data related challenges, a difficulty that often arises in comparative health assessments based on prevalences is that no group is shown to be universally healthier than the other groups across all health outcomes. For example, it has been documented that in the U.S. white elderly individuals have higher cancer prevalences than Hispanics and blacks, but that the latter groups are more likely to report having diabetes (National Research Council of the National Academics 2004). In such a situation, figuring out how to aggregate this information into a single measure without losing the benefits of prevalence-based indicators can be a challenge.¹

Our approach addresses this technical difficulty. The basic logic behind our proposal which we formalize in a set of axioms is as follows. Suppose we wish to compare two individuals, A and B, who might be suffering from cancer and cardiovascular disease. If individual A has both cancer and cardiovascular disease and individual B only has cancer, then, barring differences in the severity of the conditions, we say that individual A is in a worse health state than individual B. However, if individual A has cancer and individual B has cardiovascular disease, we say that they are not comparable. While we are focused on

¹ Fattore (2016) and Fattore and Arcagni (2018) deal with similar issues in the creation of multidimensional measures of poverty in ordinal settings.

diseases in this example, the idea can also be applied to other prevalence-based measures, such as limitations in activities of daily living or health behaviors.

Based on this reasoning, which we aggregate to population-level comparisons, we take the following three steps. First, we construct an index of comparability. In a context of multi-morbidity, this indicator captures to what extent the populations can be ranked using our axioms, and is influenced by the overlap in the profiles of conditions affecting the comparison groups. Second, we take the comparability information into account using our main indicator, the partial ordering disparity index (PODI). The PODI is an ordinal measure for population-level pairwise comparisons that ranks groups on the basis of relative prevalences. Third, we show how the PODI can be extended to generate a health ranking in multi-group comparisons. Our aim is to provide an intuitive health index that can be readily implemented using the information commonly collected in health surveys. Given the concerns that have been raised about self-assessed health measures, especially in the presence of multi-morbidity, our health indicators provide conservative estimates of the health disparities between populations that can be used to complement existing analyses. We illustrate the applicability of our analytical framework by drawing on data from the National Health Interview Survey (NHIS) to study racial health disparities. Those wishing to apply this approach in the future can refer to the user-friendly R-code and examples presented in the online appendix of this work.

The paper is organized as follows. In Sect. 2, we introduce our axioms and formalize how they can be used to provide population-level comparisons. Building on the pairwise relations discussed in Sect. 2, we present in Sect. 3 a new health measure, the partial ordering disparity index. We then derive the formal properties of the PODI and illustrate how it can be extended to compare any number of populations. In Sect. 4, we use data from the NHIS in applying the PODI to examine racial differences in health in the U.S. We discuss and compare our measures to others from the literature in Sect. 5. In Sect. 6, we conclude.

2 Definitions and Concepts

2.1 Axioms

Within the scope of our work are health comparisons based on data from health surveys, such as the National Health Interview Survey (NHIS), the Health and Retirement Study (HRS) and related surveys, and the World Health Survey (WHS). In addition to collecting data on socioeconomic variables, these surveys generally gather some information on the burden of disease at an individual level, such as self-reports on the prevalence of diseases or on limitations in daily life activities. In the analysis that follows, we assume that we have information on the prevalence of relatively common diseases, even though our approach can be implemented for any prevalence-based variable (see Sect. 5). The following axioms provide the foundation for our analysis.

Axiom 2.1 Individuals who have no condition are healthier than individuals with at least one condition.

Axiom 2.2 Individuals with the same combination of conditions are all equally (un) healthy.

Axiom 2.3 Individual j is less healthy than individual k if j has the same conditions as k and at least one additional condition.

We expect that these axioms will provide a reasonable basis for comparisons at the population level, assuming that the diseases and the limitations covered in the health survey data used accurately reflect the conditions that afflict the population in question. We wish to emphasize that our goal is to make comparisons at the population level, as we are well aware that the judgments derived from our axioms might not hold in all individual-level comparisons.

In the discussion, we explore the question of how these axioms can be extended to cover more informed assessments. The basic analytical framework is, however, unchanged by those modifications.

2.2 Partial Ordering

Axioms 1 through 3 provide a partial ordering of any two individuals based on pairwise comparisons. We define those comparisons using two relations, denoted *dominance* and *comparability*. Let $\mathcal{D} = \{D_1, \dots, D_k\}$ be the set of all conditions under study. The power set $\mathcal{C} = \mathcal{P}(\mathcal{D})$ gives the set of all possible combinations of conditions, including the case of no condition as the empty set, $\emptyset$. Elements of $\mathcal{C}$ will be denoted by C_j . Define $>$ as the non-reflexive, non-symmetric, and transitive relation of dominance and $\sim$ as the reflexive, symmetric, and transitive relation of comparability.

Definition 2.1 (Comparability) $\forall C_i, C_j \in \mathcal{C} : C_i \sim C_j$ iff $C_j \subseteq C_i$ or $C_i \subseteq C_j$.

Definition 2.2 (Dominance) $\forall C_i, C_j \in \mathcal{C} : C_i > C_j$ iff $C_j \subset C_i$.

While these relations are defined in terms of conditions, they can be easily translated into comparisons between individuals defined by condition profiles. The comparability relation captures the idea that individuals with different medical profiles might not be comparable. In turn, among comparable individuals, the dominance relation establishes a health ranking. Coming back to the example in the introduction, we can say based on Definition 2.1 that if individual A has cancer and individual B cardiovascular disease, they are not comparable. However, if individual A has both cancer and a cardiovascular disease, then A is comparable to B. Furthermore, since individual A's conditions are a strict subset of individual B's conditions, we can say that, barring differences in the severity of the conditions, individual A's health state dominates that of individual B (Definition 2.2).

The relations 2.1 and 2.2 imply a partial ordering of the elements of $\mathcal{C}$. More precisely, a strict partial order can be defined by the set $\mathcal{C}$ and the irreflexive, antisymmetric, and transitive dominance relation. It can be depicted as a directed, acyclic graph $\mathcal{G} = (N, E)$, with nodes N given by the elements of $\mathcal{C}$ and edges E given by the elements of $\mathcal{C} \times \mathcal{C}$ for which the conditions stated in Definition 2.2 apply. $\mathcal{D}$ is the maximal element of $\mathcal{G}$ and $\emptyset$ is the minimal element of $\mathcal{G}$. In a similar way a non-directed graph $\mathcal{G}'$ can be constructed that represents comparability relations (relation 2.1). An example for $\mathcal{D} = \{A, B, C\}$, given in Fig. 1, represents the dominance relation. For instance, the node $\{A, B\}$ dominates both nodes $\{A\}$ and $\{B\}$ and also the minimal element $\emptyset$ because of transitivity. In turn, it is dominated by the maximal element of the graph. A graphical representation of the comparability relation would look similar, by directed edges replaced by undirected edges.

Fig. 1 Example of partial ordering with three conditions A, B, and C. Nodes (shown as circles) represent all of the possible combinations of conditions. Edges (shown as arrows) show the dominance relation, e.g., $\{A, B\}$ dominates $\{A\}$ and $\{B\}$ and because of transitivity, also $\emptyset$. In the literature, partial orderings are often not depicted with directed edges (arrows), but with undirected edges (lines). We chose to use the former for readability

2.3 Population Level Comparisons

Having formally defined health relations at the individual level, the next logical step is to formulate a strategy to aggregate these relations for the purposes of making group comparisons. To that end, we build on the concept of expected dominance, which can be traced back to Lieberman (1976) and has also been applied to health comparisons in recent work by Herrero and Villar (2013). This intuitive measure reflects the probability that a given relation (2.1 or 2.2) can be established between any two randomly selected individuals from two different groups. Computationally, this measure corresponds to the proportion of pairwise comparisons between individuals of two groups that constitute a relation of comparability or dominance. A possible drawback of using a measure based on the “average” (expectation) is that it is blind to the presence of particular subgroups for observations for which we might observe different patterns in health disparities. However, in Sect. 3.2 we show that it is possible to decompose our health index into the contributions of different subgroups.

In this section we build on this reasoning to define two population level measures: an index of dominance and an index of comparability.

2.3.1 Population Level Comparisons: Comparability

There is no guarantee that any two groups are composed of highly comparable individuals. If relations of comparability can be established across a relatively small number of individuals only, then a health ranking might not be informative. We acknowledge this potential shortcoming in two ways. First, we construct an index of comparability following the reasoning outlined above. This index explicitly measures to what extent the disease profiles of the individuals in the two populations can be ordered using our relations. Based on the value of the index of comparability, the analyst can judge whether the results are meaningful. Second, we use the index of comparability to adjust our health disparity measure. In the following, we formalize the index of comparability; the health disparity measure is introduced later.

Let us define the index of comparability formally. Let Ω_a and Ω_b denote the sets of individuals from two groups. The proportion of observations individual ω is comparable to can be written as

$$J_i(\omega) = \sum_{C_j: C_\omega \sim C_j} P_i[C_j], \quad (2.1)$$

where C_ω is the set of conditions that afflicts ω and $P_i[C_j]$ is the proportion of individuals with conditions C_j in group Ω_i , the reference group, which can be either Ω_a or Ω_b . That is, the comparability of ω can be assessed both relative to her own group or to the comparison group. The choice of reference group is not important for our health disparity measure (see Sect. 3), but is needed to define aggregate comparability and dominance. Intuitively, comparability is between zero and 1, $0 \leq J_i(\omega) \leq 1$; i.e., in the most extreme cases individual ω might be comparable to anyone ($J_i(\omega) = 1$) or not be comparable to anyone else ($J_i(\omega) = 0$). The latter case requires that ω is not a member of Ω_i . If $\omega \in \Omega_i$, then she is at least comparable to herself, guaranteeing $J_i(\omega) > 0$.

Having defined J_i , we only need to aggregate across individuals to arrive at an aggregate measure of comparability.

Definition 2.3 The comparability index $CI_i(\Omega_j) \in (0, 1)$, with reference group Ω_i , is given by

$$CI_i(\Omega_j) = \frac{1}{n_j} \sum_{\omega \in \Omega_j} J_i(\omega).$$

The interpretation of this definition is straightforward; i.e., high CI values mean that most observations can be compared to each other, while the opposite is true when the CI is close to zero. By construction, every observation is comparable to observations with no conditions (minimal element of $\mathcal{G}$) and to observations that suffer from all conditions (maximal element of $\mathcal{G}$); and vice versa. Thus, the CI is bounded below as follows:

$$CI_i(\Omega_j) \geq P_i[\emptyset \text{ or } D](2 - P_i[\emptyset \text{ or } D]), \quad (2.2)$$

where $P_i[\emptyset \text{ or } D] = P_i[\emptyset] + P_i[D]$. For example, if $P_i[\emptyset \text{ or } D] = 0.5$, the CI is equal to or higher than 0.75, even if the observations with at least one but not all conditions are not all comparable to each other. This means that large proportions of healthy individuals in both groups will increase comparability, even if the conditions of the unhealthy individuals

differ substantially between groups. Moreover, the comparability index is symmetric in the sense that $CI_i(\Omega_j) = CI_j(\Omega_i)$.

To account for this, we can decompose the CI into two parts. The first part, $CI_i^D(\Omega_j)$, given on the right hand side of Eq. (2.2), gives the proportion of comparability due to the presence of observations with no or all conditions (see above). The second part is given by:

$$CI_i^N(\Omega_j) = \frac{1}{n_j} \sum_{\omega \in \Omega_j: C_\omega \in \mathcal{N}} J_i^N(\omega), \quad (2.3)$$

where $\mathcal{N} = C/\{\emptyset, D\}$ is the set of conditions that includes all combinations of conditions except the minimal element (healthy) and the maximal element (all conditions) and

$$J_i^N(\omega) = \sum_{C_j: C_\omega \sim C_j} P_i[C_j] \text{ for } C_j, \quad C_\omega \in \mathcal{N}. \quad (2.4)$$

This gives the contribution to the overall comparability of individuals with at least one but not all conditions; in other words, to the comparability among the sick who might not be comparable. Before turning to dominance, we wish to make a final important remark regarding the interpretation of comparability. All of our measures are aggregations of individual-level comparisons. For this reason, even if two populations have identical distributions of prevalences at the population level, they might have a low comparability index if the individuals constituting the groups are not comparable. Ultimately, this can be the case because our measures of comparability are all based on the concept of dominance, and not on dissimilarity. One individual is not comparable to another individual, not only because each has a different affliction, but because it is not possible to establish a relation of dominance between them. The clearest example of such a situation is a comparison between a completely healthy individual and an individual suffering from several conditions. While these individuals may satisfy the comparability criteria, they clearly have very different medical profiles. At the same time, as a necessary but not sufficient condition for achieving high levels of comparability, populations have to suffer from similar diseases.

2.3.2 Population Level Comparisons: Dominance

The procedure we used to count the proportion of comparable individuals can also be applied when assessing the proportion of individuals linked by a relation of dominance. In addition, we build on this reasoning and include a comparability correction: we propose weighting the dominance relations not by the total possible pairwise comparisons, but rather by the total comparable relations only. In essence, our axioms provide us with tools that can be used to establish rankings among the comparable, while staying neutral among the non-comparable. For this reason, while we still report information on the total proportion of comparable relations, we exclude those individuals for whom we cannot establish an ordering from dominance measures.

Let Ω_a and Ω_b denote the sets of individuals from the two groups to be compared; let $P_i[C_j]$ be the proportion of individuals with conditions C_j in group Ω_i . For individual ω with conditions C_ω the proportion of dominated individuals is given by:

$$I_i(\omega) = \sum_{C_j: C_\omega \succ C_j} P_i[C_j]. \quad (2.5)$$

For a given reference group, $I(\omega)$ can be computed for each individual $\omega \in \Omega_a, \Omega_b$.

Note that since $J_i(\omega) \geq I_i(\omega)$ the expectation of $I_i(\omega)$ is bounded by comparability. To account for this, we count the number of relations of dominance over the comparable pairwise relations instead of the total. This correction defines the comparability weighted health dominance index.

Definition 2.4 The comparability weighted health dominance index $\text{CHDI}_i(\Omega_j) \in [0, 1]$, with reference group Ω_i , is given by

$$\text{CHDI}_i(\Omega_j) = \frac{1}{n_j} \sum_{\omega \in \Omega_j} \frac{I_i(\omega)}{J_i(\omega)}.$$

Note that the CHDI is not defined when there are uncomparable sets of conditions, i.e., when at least one member in Ω_j is not comparable to any member of Ω_i , as this leads to division by zero because $J_i = 0$.

There are two issues that remain unresolved. First, the values of the comparability index and of the CHDI are dependent on the choice of reference group. Second and more importantly, we require a way to compare health across multiple groups since, as we show in the next section, dominance measures are not always transitive. That is, such measures do not necessarily lead to a unique ranking of populations. Our proposed health ranking indicator addresses both concerns.

3 The PODI Index

Having formally defined the relations of dominance and comparability at the population level, we propose the partial ordering disparity index (PODI). The PODI is constructed through a simple modification of the CHDI that resolves the issue of having to choose a reference population. In the first part of this section, we present the PODI and show that, despite its simplicity, the index has a number of desirable properties: i.e., it is not altered by the sample size or the choice of reference group, and it properly reflects the health improvements within groups. Then, in the remainder of the section, we illustrate how the PODI can be used to establish a health ranking between any number of groups. In conclusion, we demonstrate how it is possible to capture the essence of the axioms we presented in the first part of the paper in a health measure capable of ranking across multiple groups or populations.

3.1 Definition

We define the PODI between two groups as the difference between the CHDIs resulting from alternating the reference group. This essentially captures the difference in expected domination.

Definition 3.1 For two groups Ω_a and Ω_b the $\text{PODI} \in [-1, 1]$ is defined as

$$P_{ab} = \text{CHDI}_a(\Omega_b) - \text{CHDI}_b(\Omega_a).$$

As noted above, the CHDI captures dominance among the comparable individuals. Thus, a value of 1 of the PODI means that among the comparable individuals every

member of group Ω_b is less healthy than every member of group Ω_a ; the reverse is true when $P_{ab} = -1$. A value of zero implies that both groups are equally (un)healthy. Intermediary values can also be interpreted directly; values above zero indicate the disadvantage of Ω_b relative to Ω_a , while the opposite is the case for values below zero. Intuitively, the PODI measures the relative expected dominance of one group over the other. For example, $P_{ab} = 0.3$ means that on average, members of Ω_b dominate 30% more members of Ω_a than members of Ω_a dominate members of Ω_b . A value of -0.3 again indicates a reversed situation. This also means that a given value x of the PODI can be due to many different combinations of values of the CHDI.

3.2 Properties

The PODI has a number of interesting properties. The choice of the reference group, the sample size, and the definition of the pairwise relations do not alter the results in unexpected ways.

Property 3.1 (Semi-symmetry) *If the distributions of conditions are changed for groups Ω_a and Ω_b , the sign of the PODI is reversed, i.e. $P_{ab} = -P_{ba}$, otherwise it remains unchanged.*

Property 3.2 (Sample size independence) *The PODI does not depend on sample size, only on relative frequencies. That is, if n_a and n_b denote the sizes of populations Ω_a and Ω_b and n_c and n_d are the sizes of two other populations Ω_c and Ω_d for which $n_a \neq n_c$ and $n_b \neq n_d$ hold, but the distribution of conditions is the same for Ω_a and Ω_c as well as for Ω_b and Ω_d , then $P_{ab} = P_{cd}$.*

Property 3.3 (Mirror property) *If the dominance relation as given in Definition 2.2 is reversed, i.e. $\forall C_i, C_j \in C : C_i \leq C_j$ iff $C_j < C_i$, calculating the “reversed” PODI yields $1 - P_{ab}$.*

Furthermore, the PODI is well behaved with respect to changes in the prevalence of conditions that result in an improvement in the health of one of the groups. We consider two changes that unambiguously improve the health of a group and demonstrate that those changes also improve the PODI. First, let one condition be removed from a member of group Ω_a ; call this a cure. The second change consists of adding an individual with no conditions to group Ω_a ; we call this a group improving addition. Let Ω_{a_c} be the group resulting from curing one member of a , and Ω_{a_A} be the group resulting from adding a healthy individual to group a . Then, the following properties hold (proofs are given in the “[Appendix](#)”).

Property 3.4 (Cures) *Cures do not decrease the PODI, i.e. $P_{ab} \leq P_{a_c b}$.*

Property 3.5 (Group improving additions) *Group improving additions do not decrease the PODI, i.e. $P_{ab} \leq P_{a_A b}$.*

Finally, the PODI between two populations can be decomposed into the contributions of distinct subpopulations. For example, suppose we are interested in examining the relative health of the males and the females in a given country. We might suspect that these differences are not constant across, for example, socioeconomic groups. This hypothesis can

be tested directly since, as we show, the PODI is the weighted sum of the relative health across subpopulations.

Formally, suppose populations Ω_a and Ω_b can each be divided in S subgroups, denoted with the superindex s . For each subgroup s the PODI can be calculated as $P_{ab}^s = \text{CHDI}_a^s(\Omega_b^s) - \text{CHDI}_b^s(\Omega_a^s)$. Then, the aggregate PODI can be decomposed into weighted contributions of each subgroup comparison and an additional cross-group comparison, denoted by P_{ab}^{-s} .

Property 3.6 (Decomposability) *The PODI can be decomposed as*

$$P_{ab} = \sum_{s=1}^S w^s P_{ab}^s + P_{ab}^{-s}, \quad (3.1)$$

where w_s is a weight given by $\frac{n_a^s n_b^s}{n_a n_b}$. P_{ab}^s equals

$$\frac{1}{n_b^s} \sum_{\omega \in \Omega_b^s} \frac{I_a^s(\omega)}{J_a^s(\omega)} - \frac{1}{n_a^s} \sum_{\omega \in \Omega_a^s} \frac{I_b^s(\omega)}{J_b^s(\omega)} \quad (3.2)$$

with $I_a^s(\omega)$ defined as $\sum_{C_j: C_\omega > C_j} P_a^s[C_j]$, where $P_a^s[C_j]$ is the distribution of conditions in Ω_a^s . $J_a^s(\omega)$, $I_b^s(\omega)$, and $J_b^s(\omega)$ are defined correspondingly. P_{ab}^{-s} is given by

$$\frac{1}{n_b} \sum_{\omega \in \Omega_b^s} \frac{I_a^{-s}(\omega)}{J_a^{-s}(\omega)} \frac{1 - n_a^s}{n_a} - \frac{1}{n_a} \sum_{\omega \in \Omega_a^s} \frac{I_b^{-s}(\omega)}{J_b^{-s}(\omega)} \frac{1 - n_b^s}{n_b} \quad (3.3)$$

with $I_a^{-s}(\omega) = \sum_{C_j: C_\omega > C_j} P_a^{-s}[C_j]$ and $P_a^{-s}[C_j]$ is the distribution of conditions in Ω_a/Ω_a^s . $J_a^{-s}(\omega)$, $I_b^{-s}(\omega)$, and $J_b^{-s}(\omega)$ are defined correspondingly.

A proof is given in the “[Appendix](#)”. This property shows that a decomposition by subgroups as described above is relatively straightforward.

3.3 A Health Ranking

In applying any measure of health, we often aim to draw conclusions based on comparisons that involve more than two groups. Unfortunately, the pairwise measures defined in the previous sections can not be applied directly to these scenarios since they are not always transitive.

Property 3.7 (Non-transitivity) *The PODI is non-transitive in the sense that if $P_{ab} > 0$ and $P_{bc} > 0$ (or $P_{ab} < 0$ and $P_{bc} < 0$) then $P_{ac} \leq 0$.*

In our case, this issue can arise when there is a high degree of non-comparability across groups (an example is provided in the “[Appendix](#)”).

This well-known shortcoming has been addressed in the literature on the optimal aggregation of pairwise comparisons (e.g. Gavalec et al. 2015), and on pairwise comparisons of

health (Herrero and Villar 2013). A common strategy for overcoming non-transitivity is to summarize all pairwise comparisons P_{ij} for each group i with all other groups j into a single index I_i . By construction, this index is transitive and thus provides a health ranking across groups.

Here we follow the approach developed in Crawford and Williams (1985) and Barzilai and Golany (1990). Intuitively, the strategy is to obtain individual group indexes that are as close as possible to the values of each of the possible pairwise comparisons. Since this cannot be fully achieved, the method specifies a penalty function consisting of taking the square of the differences between P_{ij} and the values for each group I_i .

Let $\Omega_1, \Omega_2, \dots, \Omega_G$ be the groups for which we know P_{ij} and want to derive I_i . The individual I_i are normalized to sum to zero, i.e., $\sum I_i = 0$. Moreover, the individual index I_i to be as close as possible to the pairwise PODI between any pair of populations. That is, that pairwise differences of index values should mimic pairwise comparisons between populations given by P_{ij} as closely as possible. To achieve this, we minimize

$$\min_{I_i \in \{I_1, \dots, I_S\}} \sum_{i=1}^G \sum_{j=1}^G [P_{ij} - (I_i - I_j)]^2, \quad (3.4)$$

given the constraint that

$$\sum_i I_i = 0. \quad (3.5)$$

An optimal solution to the problem is given by:

$$I_i = \frac{1}{P} \sum_{j=1}^P P_{ij}, \quad (3.6)$$

which is simply the mean of all pairwise comparisons for group i .

This solution guarantees that higher values of I_i imply that a given group i has a relatively low health status. Furthermore, Barzilai and Golany (1990) prove that the solution is unique; i.e., that there are no other optimal solutions to (3.4), and it does not depend on the ordering of groups; i.e., that relabeling groups does not change the result. Note that the additive restriction (3.5) is required for a unique solution, but that the choice of normalization does not affect the ranking. In the next section we apply our concepts to a comparison of health across races in the U.S.

4 Empirical Application

We use our health indexes to study racial health differences in the US. Drawing on data from the National Health Interview Survey (NHIS) for the year 2014, we study the relative prevalence of 15 conditions across four races for 3570 individuals. The comparison is narrowed to ages 60 to 65 to control for the differences in age distributions across races. The objective of this exercise is to illustrate that while simple comparisons of prevalences across groups are appealing, they are often not sufficient to rank health across populations or groups.

Table 1 Prevalence by race

Diseases	White (%)	Asian (%)	Black (%)	Hispanic (%)	All races (%)
Hypertension	48.1	49.1	69.6	52.9	51.6
High cholesterol	48.8	45.6	49.1	46.4	48.5
Coronary heart disease	7.0	4.3	8.0	9.1	7.2
Angina pectoris	2.9	1.2	3.4	3.0	2.9
Heart attack	6.1	4.9	5.2	6.1	5.9
Heart condition/disease	11.9	7.4	12.1	7.2	11.2
Stroke	4.2	1.8	9.1	6.3	5.0
Emphysema	3.9	0.6	4.0	3.6	3.7
COPD	7.7	3.1	7.4	3.6	7.1
Asthma	12.7	12.3	17.7	11.5	13.3
Ulcer	10.0	6.8	9.9	11.5	10.0
Cancer	17.4	5.5	9.5	10.7	15.1
Diabetes	16.0	17.8	29.4	29.1	19.3
Arthritis	40.8	25.9	47.2	35.7	40.5
Chronic liver condition	2.7	2.5	2.8	3.1	2.7

4.1 Descriptive Statistics

We report the prevalences of the full set of 15 diseases in the NHIS for our age and racial groups (Table 1). Prevalences are measured by survey questions that ask respondents whether they have ever been told by a doctor that they suffer from a given condition. Because they are relatively small in number, the individuals with missing values are removed from the sample.² In addition, we restrict our exercise to the individuals who reported belonging to one of the four main racial categories, and exclude those who reported belonging to the other category, as this category is difficult to interpret. The data should be read with some caution since the narrow age group results in small samples for some racial groups. For example, the Asian group is composed of 163 individuals. As a consequence, the number of cases of some of the less prevalent diseases might be subject to small sample bias.

A first look at the relative prevalences across the races reveals that some conditions have a much steeper race gradient than others. For example, black Americans have a much higher prevalence of hypertension (69.6%) than other races (at around 50%). In contrast, the differences in the prevalence of heart attacks across races appear to be much smaller (between 4.9% for Asian individuals, and 6.1% for Hispanic and white individuals). Nevertheless, the table shows a wide range of prevalences across races for each disease. Given that no racial group has unambiguously lower prevalences than the other groups across all conditions, it is not possible to establish a relative health ordering without further analysis.

In addition, Table 2 highlights the empirical relevance of multi-morbidity. Only 15.3% of the sample reported having none of the 15 diseases, while 20.8% indicated that they have one condition. The remaining 63.9% of the individuals across all races reported suffering

² There are a total of 153 missing values and no racial group has a disproportionate share of missing values.

Table 2 Number of diseases per person by race

	0 (%)	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	6 (%)	Total (%)
White	16.0	21.2	22.9	16.0	10.7	5.9	7.4	100.0
Asian	21.3	25.8	22.6	14.8	7.7	4.5	3.2	100.0
Black	9.2	17.4	21.9	18.4	16.6	8.0	8.4	100.0
Hispanic	16.8	20.7	23.3	16.5	7.7	7.1	8.0	100.0
All races	15.3	20.8	22.7	16.3	11.1	6.3	7.6	100.0

Table 3 Most prevalent combinations of diseases by race

Race	Hypertension and cholesterol (%)	Hypertension and arthritis (%)	Cholesterol and arthritis (%)	Rest (%)
White	22.1	19.3	12.7	45.8
Asian	28.6	11.4	14.3	45.7
Black	30.8	5.6	13.1	50.5
Hispanic	15.9	14.6	13.4	56.1
All races	23.0	16.5	12.9	47.7

from at least two conditions. While the prevalence of multi-morbidity differs across races, in each of the racial groups more than half the individuals said they have at least two conditions, and a high proportion indicated that they have several more. A further difference across that races is that among the individuals suffering from multi-morbidity, the most frequent combinations of conditions vary in prevalence as well. For example, among those individuals with two conditions, the combination of hypertension and cholesterol is twice as prevalent among blacks (30.8%) as among Hispanics (15.9%) (Table 3).

Our health measures are designed to provide a health ranking for precisely this type of complex information on prevalences. First, we assess the comparability of the group-specific profiles of diseases. Then, once we have established whether a prevalence-based comparison is reasonable, the PODI provides a health ranking.

4.2 Health Measures

4.2.1 Comparability

Overall, the comparability index across races (Table 4, panel A) ranges from .51 (between blacks and whites) to .59 (between Asians and Hispanics). This means that it is possible to establish relations of comparability for over half of the possible pairwise comparisons across individuals of different races. In this example, we have purposely kept diseases at the lowest possible form of aggregation, which makes comparability more difficult. Hence, we treat this as a lower bound to the comparability levels. To illustrate the impact of the level of aggregation in conditions, we have calculated the comparability index after merging all of the circulatory diseases into one category (following the IC-10 category definitions). Grouping circulatory conditions has a very

Table 4 Comparability

	White	Black	Asian	Hispanic
<i>(a) Comparability disaggregated</i>				
White	0.54	0.52	0.58	0.54
Black	0.52	0.51	0.58	0.53
Asian	0.58	0.58	0.63	0.59
Hispanic	0.54	0.53	0.59	0.55
<i>(b) Comparability aggregated</i>				
White	0.71	0.70	0.75	0.70
Black	0.70	0.69	0.75	0.69
Asian	0.75	0.75	0.79	0.75
Hispanic	0.70	0.69	0.75	0.70

Table 5 PODI

Ranking		White	Black	Asian	Hispanic
<i>(a) PODI disaggregated</i>					
3rd	White	–	0.15	–0.16	–0.02
4th	Black	–0.15	–	–0.31	–0.16
1st	Asian	0.16	0.31	–	0.14
2nd	Hispanic	0.02	0.16	–0.14	–
<i>(b) PODI aggregated</i>					
2nd	White	–	0.12	–0.17	0.01
4th	Black	–0.12	–	–0.29	–0.10
1st	Asian	0.17	0.29	–	0.18
3rd	Hispanic	–0.01	0.10	–0.18	–

noticeable impact on comparability, raising the lowest comparability index to .69 and the highest to .79 (Table 4, panel B). Results on decomposed comparability for the disaggregated example are shown in the “Appendix” (Tables 8 and 9).

4.2.2 PODI and Dominance

Table 5 shows both of the pairwise PODI indexes (3.1) across races and the summary index (3.6) that aggregates the pairwise comparisons; results on the underlying values of the CHDI are given in the “Appendix” as an example for the disaggregated case. We include both the PODI for the full set of conditions and the PODI for the aggregation of circulatory diseases. In this particular case, we find that the pairwise PODIs satisfy transitivity, but this need not be the case in the other scenarios. Across both scenarios, our results indicate that the Asian group is the healthiest, and the black group is the least healthy. However, the intermediate positions depend on the modeling of circulatory diseases. In our preferred model, with a single category for circulatory diseases, the white group is ranked higher than the Hispanic group. The opposite outcome is found when we use multiple circulatory disease categories; and, hence, when comparability is low.

4.2.3 Comments

The results of our application are generally in line with existing findings on racial/ethnic disparities in health in the US (for a recent overview, see Braveman et al. 2010). It is well-established that morbidity levels are highest among the black population (for a recent reference, see Ward 2013). In addition, the results highlight the importance of taking comparability into consideration. The Hispanic group is ranked as healthier than the white group in the low comparability case only. This finding corresponds with recent research indicating that the so-called Hispanic paradox³ might not apply to morbidity. However, the prevalences found among the Asian population should be viewed with care given the relatively small sample size for this group. For an in-depth study on racial differences in health among the elderly in the U.S. that includes such measures, see, for example, Hummer et al. (2004).

5 Discussion

5.1 Methodological Considerations

In this paper we have explored a specific application of a partial order approach with condition prevalence as a more robust objective measure of health than self-reported assessments. Our current implementation focused on disease prevalences allows us to readily compute our health index in most health surveys. However, it would be possible to extend our analysis to include expert knowledge on the relative severity of different diseases. Such a refinement would entail modifying the definitions of the relations of dominance and comparability to include additional criteria, but the analytical framework would remain unchanged. Other extensions, and applications to different kinds of data can be considered in our framework; the theoretical backbone of our measures can be adapted to those scenarios. For example, we could apply the PODI as an alternative index to aggregate the information of health surveys on multiple domains of health, such as the EQ-5-D (Devlin and Brooks 2017). In that vein, the PODI can be applied to the types of data that are used to construct other more general multidimensional indexes of well-being that are not restricted to health. Examples include, but are not limited to, the Better Life Index developed by the Organization for Economic Co-operation and Development (OECD), or the Human Development Index sponsored by the United Nations (UN). The analytical structure of our indicators can support ordinal variables with multiple categories, as these applications would require, though modifications of our R functions would be required.⁴

In addition to the issues related to data collection (diagnosis bias and reporting bias) already discussed, a potential drawback of the simple criteria on which our approach is based is that they could result in low comparability; i.e., in health rankings that only capture reliably a fraction of the population. Our recommendation for potential users is to avoid including too many individual conditions or dimensions in the analysis. Aggregations

³ In the U.S., Hispanic individuals tend to have a lower socioeconomic status than white individuals. However, some studies have reported a relative mortality advantage for the Hispanic group. This finding has been called the Hispanic paradox (e.g. Abraido-Lanza et al. 1999).

⁴ We provide R functions that currently only accept binary measures. Alternatively, the different variables could be collapsed into binary variables.

across typologies of diseases are a straightforward way to increase comparability. Still, if the focus of the analysis is not on individual conditions, using coarser definitions for disease profiles might result in only a small loss of information. In other contexts, limiting or aggregating the number of ordinal categories per dimension could be an option. These are just two examples, and ultimately the best approach is clearly dependent on the application.

5.2 Comparison to Other Approaches

There exists a variety of alternative measures of population health and well-being which emphasize different issues. In order to clarify the contribution of the measure we have developed in this article, we discuss our proposal with regard to three closely related literatures. First, we discuss our treatment of multi-morbidity compared to the clinical literature, which explores consequences of disease combinations on specific health outcomes. Second, we relate our approach to existing research on the measurement of multi-dimensional well-being that focuses on the aggregation problem, most frequently using weighting approaches. Finally, we review alternative analysis that also utilize partial order approaches to generate rankings in multi-attribute situations as a response to the criticisms levied on weighting schemes.

5.2.1 Clinical and Public Health

Our focus on disease prevalences is perhaps closest to clinical and public health research. The clinical literature on multi-morbidity has explored measures that evaluate combinations on diseases based on their influence on highly specific health outcomes of interest (de Groot et al. 2003). For example, the Charlson index (Charlson et al. 1987) is a widely used health indicator which is computed as the weighted combination of an array of conditions, and it is validated by its capacity to predict the future mortality of the patients. Much like the Carlson index, many other clinical indicators are designed to assess multi-morbidity in specific clinical applications. For example, the Hallstrom index (Hallstrom et al. 1996) focuses on the prediction of cardiac episodes, and the Hurwitz index (Hurwitz and Morgenstern 1997) assesses how multi-morbidity influences the healthcare choices of patients with back pain.

A limitation of these measures is that the strong influence of the application on the design of the indexes can result in a poor fit when used to study other issues. Brilleman et al. (2014) evaluate the capacity of different multi-morbidity indexes to predict patient health care costs, and conclude that the Carlson index is over-performed by simple disease counts. Furthermore, these methods also have strong data requirements, such as mortality follow-ups, that go beyond the information collected in most national health surveys. In some cases, the methods need to be calibrated to each specific population, which results in further data demands.⁵

Perhaps in an effort to avoid the specificity of the multi-morbidity measures in clinical research, the public health literature has often opted for simpler approaches. Some prominent studies rely on disease-by-disease prevalence comparisons without further elaboration (Banks et al. 2006). Other studies, more focused on multi-morbidity, use disease counts as

⁵ For example, the Charlson index is calibrated using a set of weights that depend on the effects of conditions on mortality and may vary for each population.

their primary index (Marengoni et al. 2008; Ward 2013). In the context of these studies, the PODI allows us to establish a more systematic approach to population level health assessments. In addition, we can generate measures of comparability that provide meaningful summaries of the overlap of the condition profiles between the populations analyzed. This is important given that, as we have shown in our empirical application and is recognized in the clinical literature, multi-morbidity contains a wide diversity of complex disease profiles that often have a socioeconomic gradient (Schiøtz et al. 2017).

5.2.2 Health Economics

In the field of health economics there have been several proposals to generate indexes that aggregate across multiple dimensions of health on the basis of self reported health data. A well known example of this approach is the EQ-5D methodology, developed by the EuroQol Group (Devlin and Brooks 2017).⁶ The basis of the EQ-5D approach is a survey in which individuals are asked to self assess their health state in a variety of dimensions, ranging from the pure physical (pain), to limitations (or lack of) on their daily life activities as a consequence of their health. These measures are then aggregated into a single index.

Setting aside the implementation requirements for now, the tradeoff of utilizing such measures is the introduction of possibly contentious assumptions. First, these indexes rely on self-assessed measures, and as such have been shown to be vulnerable to reporting heterogeneity, a form of bias. This issue occurs when, for a similar objective state of health, individuals report systematically different self assessments depending on their socioeconomic characteristics, or country of pertinence (Bago d'Uva et al. 2008). Second, aggregating the self assessed limitation into a single index requires attributing preference weights or relative importances to the different dimensions of limitation. As the developers of the EQ5-D acknowledge (Devlin et al. 2018), there remains considerable debate on the methods used to evaluate preferences over certain health states. In addition, using preference-based weights might not be desirable in all applications; in some cases the analyst might prefer comparisons based purely on objective health differences, and not confounded by individuals' valuations. Last but not least, such summary measures and questionnaires are not always implemented, whereas disease and limitation prevalences appear in most health national surveys.

5.2.3 Partial Orders

An alternative to weighting schemes consists in a variety of approaches that rely on partial orders. We identify two main existing distinct approaches that, like us, use partial order notions as an alternative to the more traditional weighting approach to deal with multiple attributes. The research by Fattore, Maggino and coauthors (henceforth referred to as "LE") uses the concept of linear extensions as a natural partial order ranking (Fattore et al. 2011; Fattore and Maggino 2015; Fattore and Arcagni 2018). A different approach, developed in a series of papers starting with Arndt et al. (2012) (henceforth referred to as "FOD"), is based on an application of the multidimensional version of first order stochastic dominance (Hussain et al. 2016; Permanyer and Hussain 2018).

⁶ There are others, such as the 36-Item Short Form Health Survey (SF-36) proposed by the RAND (Stewart et al. 1992).

5.2.4 Linear Extensions

The work by LE builds on the counting method developed by Alkire and Foster (2011), which classifies individuals as poor if they are below the poverty line in more than a given number of well-being dimensions. The authors point out that, from a partial order perspective, it is not clear that all profiles below and above the poverty line are comparable.⁷ Instead of simple counts, the authors propose establishing a set of deprived profiles, and proceed to compare each individual profile to them using partial order theory. However, note that while LE focus on the measurement of poverty, existing work in other fields has similar strategies to obtain pure orderings (Fattore and Bruggemann 2017).

There is a main difference between our approach and that in the literature that uses linear extensions. Primarily, we differ in the treatment of non-comparable profiles. Instead of using the relation for non-comparability, the authors opt for a fuzzy approach to classifying profiles. The main thrust of their and related contributions consists of exploiting the concept of linear extensions, and the fuzziness arises from the fact that one can obtain multiple linear extensions from a partially ordered set. Succinctly put, a linear extension is a linear order such that it respects the ordering of the original set of profiles; in scenarios with non-comparable profiles, a partial order has multiple linear rankings. The approach in Bruggemann and Carlsen (2011) and others (e.g., Morton et al. 2009) consists in dealing with the multiplicity of orders condensing the information in summary statistics computed from all the linear extensions. Some methods opt to eliminate the ambiguity from non-comparable profiles altogether, by, for example, taking the average position across linear orders as its final ranking. Others, like LE, report the proportion of linear orders that classify a given profile as poor, hence preserving the non conclusiveness of the linear orders. In our proposal, we have opted instead for excluding non-comparable relations from our rankings, and thus focus only on those cases for which an unambiguous order can be established. Bear in mind that this does not equate to completely eliminating certain profiles from the analysis; rather, the more comparable profiles will have a greater importance in the overall assessment. The robustness of the PODI ranking to the treatment given to unclear comparisons can be evaluated on the basis of the comparability index.

5.2.5 First Order Stochastic Dominance

The FOD approach is another methodology that imposes strong requirements on the types of populations that can be compared. The contribution by the scholars working with FOD applied to multidimensional indexes consists in operationalizing the multidimensional extension of the concept of first order stochastic dominance in a context with many individuals and attributes. While FOD based evaluations provide uncontroversial orders, they suffer from two main limitations. The first is a technical limitation. While some progress has been made in this vein, there remain feasibility constraints on the number of dimensions for which FOD tests can be computed. Second, the main drawback of the FOD approach is that the conditions needed to establish dominance of a distribution over an other are so

⁷ To see this, consider an example with two countries, 1 and 2, and three attributes, and define poverty to require falling below the poverty line in at least two attributes. Suppose country 1 is poor in dimensions one and three, while country 2 only in dimension two. Then, from the perspective of the (partial) order relations, we cannot establish with clarity that the first country is worse off than the second, even though one country is classified as poor and the other is not.

Table 6 Distribution of prevalences across three populations

Conditions	Pop.1 (%)	Pop.2 (%)	Pop.3 (%)
A&B	25	30	15
A	25	10	25
B	25	10	25
Healthy	25	50	35

Table 7 PODI and FOD assessments

Ranking		Pop.1	Pop.2	Pop.3
<i>(a) PODI</i>				
3rd	Pop.1	0	-0.161	-0.167
1st	Pop.2	0.161	0	0.034
2nd	Pop.3	0.167	-0.034	0
<i>(b) FOD</i>				
2nd	Pop.1	–	n.c	3FOD1
–	Pop.2	n.c	–	n.c
1st	Pop.3	3FOD1	n.c.	–

n.c. stands for not comparable

demanding that frequently it does not allow us to judge between distributions in applications. Permanyer and Hussain (2018) provide examples and simulations that illustrate this point; essentially, FOD applicability diminishes considerably when the number of attributes increases, particularly when the attributes are not highly correlated. In the application provided in the paper, with multi-attribute cross country evaluations, 47% of the comparisons can be ranked according to FOD for three dimensions, dropping to 31% and 10% for five and ten dimensions respectively.

Compared to FOD based analysis, our proposal allows the applied researcher to compare more frequently populations across multiple dimensions, at the cost of relaxing the requisites to establish an order. The basic intuition for our approach, which we have refined and formalized, is that an order can be established if, in expectation, an individual from a given population can be deemed healthier than that of another. To illustrate the differences between the two approaches, consider the following example from Arndt et al. (2012). Table 6 describes the prevalence of two conditions across three populations, and Table 7 the orders according to our method (panel A), and following the FOD approach (panel B). Comparability is high in this example ($\geq .85$), and thus our method covers a very large fraction of the possible relations across populations. The PODI index ranks all populations, compared to the FOD approach that is restricted to populations 1 and 3; furthermore, note that our methods coincides with the FOD for those particular populations.

6 Conclusion

We propose a new indicator, the PODI, based on the relative prevalence of conditions across populations. Although no statistic can provide an universal solution for the methodological issues that arise when assessing health disparities, our method specifically

addresses two of the main difficulties associated with health comparisons. That is, we provide a measure that is both robust to self-assessment biases, and that takes into consideration the differences in the types of diseases afflicting the comparison groups. We illustrate how our simple framework can be used to analyze complex data on disease prevalences by applying it to current health disparities across racial groups in the U.S.

Acknowledgements Open access funding provided by Max Planck Society. The authors thank Mikko Myrskylä, Jo Mhairi Hale, Adeline Lo, and three anonymous referees for their comments and helpful suggestions. Manja Janowski provided excellent research assistance. All errors are our own.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix 1: Proofs

Properties 3.1, 3.2, and 3.3 follow directly from the definition of the PODI. The proof of Property 3.5 follows from the proof of Property 3.4 which is given below.

Proof of Property 3.4 A simple proof of Property 3.4 uses the concept of paths from the maximal to the minimal node, also called chain in the theory of partially ordered sets. A path is any sequence of nodes and edges from D to $\emptyset$ with all nodes and edges being unique and the path following the direction of the edges. For instance, in Fig. 1 going from D to $\{A, B\}$ to $\{A\}$ to $\emptyset$ is a path, while $D, \{A, B\}, \{A\}, \{A, C\}, \{C\}, \emptyset$ is not. Using this concept it is easy to show that $\text{CHDI}_{a_c}(\Omega_b) \geq \text{CHDI}_a(\Omega_b)$ and $\text{CHDI}_b(\Omega_a) \geq \text{CHDI}_b(\Omega_{a_c})$ hold, which proves the above property. Let ω_c be the cured individual from population a . Now we can distinguish four cases:

- For individuals from group Ω_b who are not on the same path as ω_c both before and after her cure the contribution to the CHDI will not change.
- For individuals from Ω_b who are on the same path before and after the cure comparability will not change. As dominance may increase, but cannot decrease, the contribution of dominance to the CHDI will stay constant or increase.
- For individuals from Ω_b who are on the same path as ω_c before the cure but not afterwards, comparability will decrease. However, dominance will not change, as a necessary requirement for ω_c to leave the shared path is that she has more conditions than the individuals from Ω_b we are concerned with.
- For individuals from Ω_b who are not on the same path as ω_c before the cure but are on the same path afterward, comparability will increase. Moreover, dominance will also increase. This happens because in order to move on the same path, ω_c must have the same number of conditions as the members of Ω_b to whom this situation applied before the cure. After the cure they are on the same path and ω_c has one less condition and is therefore dominated. Thus, for each member of Ω_b for whom this situation applies the contribution to the CHDI is $\frac{I_a(\omega)+1/n_a}{J_a(\omega)+1/n_a}$, which is bigger than or equal to $\frac{I_a(\omega)}{J_a(\omega)}$ because

$$I_a(\omega) \leq J_a(\omega).$$

This proves that $\text{CHDI}_{a_c}(\Omega_b) \geq \text{CHDI}_a(\Omega_b)$ holds. $\text{CHDI}_b(\Omega_a) \geq \text{CHDI}_b(\Omega_{a_c})$ is trivial, as only ω_c 's contribution to the CHDI is affected and can only decrease. $\square$

Proof of Property 3.6 It suffices to show that the CHDI can be decomposed by groups. The decomposition of the PODI is just the difference between the decompositions. The CHDI is given by

$$\begin{aligned}\text{CHDI}_a(b) &= \frac{1}{n_b} \sum_{\omega \in \Omega_b} \frac{I_a(\omega)}{J_a(\omega)} \\ &= \frac{1}{n_b} \sum_{\omega \in \Omega_b} \Pr(\omega > \omega_a | \omega \sim \omega_a),\end{aligned}$$

where $\Pr(\omega > \omega_a | \omega \sim \omega_a)$ is the probability that a member ω_a of Ω_a dominates ω from Ω_b , conditional on being comparable. It can be decomposed as

$$\text{CHDI}_a(b) = \sum_{s=1}^S \frac{n_a^s n_b^s}{n_a n_b} \frac{1}{n_b^s} \sum_{\omega \in \Omega_b^s} \frac{I_a^s(\omega)}{J_a^s(\omega)} + \frac{1}{n_b} \sum_{\omega \in \Omega_b^s} \frac{I_a^{-s}(\omega)}{J_a^{-s}(\omega)} \frac{1 - n_a^s}{n_a}.$$

Rearranging gives

$$\begin{aligned}\text{CHDI}_a(b) &= \frac{1}{n_b} \sum_{s=1}^S \sum_{\omega \in \Omega_b^s} \Pr(\omega > \omega_a | \omega \sim \omega_a \text{ and } \omega_a \in \Omega_a^s) \Pr(\omega_a \in \Omega_a^s) \\ &\quad + \frac{1}{n_b} \sum_{s=1}^S \sum_{\omega \in \Omega_b^s} \Pr(\omega > \omega_a | \omega \sim \omega_a \text{ and } \omega_a \notin \Omega_a^s) \Pr(\omega_a \notin \Omega_a^s),\end{aligned}$$

which equals $\frac{1}{n_b} \sum_{\omega \in \Omega_b} \Pr(\omega > \omega_a | \omega \sim \omega_a)$ and completes the proof. $\square$

Proof of Property 3.7 We provide a non-transitive example. Suppose we have three groups, Ω_a , Ω_b , and Ω_c , and three conditions, A , B , and C . Assume that in group Ω_a 30% of individuals have no condition and 70% have condition B . In group Ω_b 30% have no condition, 50% have condition A , and 20% have both condition A and B . In group Ω_c 50% have no condition and 50% have both condition A and C . The values of the PODI between the populations

described in this example are $P_{ab} = 0.22$, $P_{bc} = 0.05$; thus $P_{ac} = -0.2$, and thus the PODI is not transitive. $\square$

Appendix 2: Additional Tables

See Tables 8 and 9.

Table 8 Decomposed comparability

	White	Black	Asian	Hispanic
<i>(a) Comparability by definition, disaggregated</i>				
White	0.29	0.24	0.34	0.30
Black	0.24	0.18	0.29	0.24
Asian	0.34	0.29	0.38	0.34
Hispanic	0.30	0.24	0.34	0.31
<i>(b) Comparability among the sick (excluding individuals with all conditions), disaggregated</i>				
White	0.30	0.31	0.31	0.29
Black	0.33	0.37	0.37	0.34
Asian	0.29	0.32	0.31	0.30
Hispanic	0.29	0.31	0.31	0.30

Table 9 CHDI disaggregated

	White	Black	Asian	Hispanic
White	0.58	0.66	0.50	0.58
Black	0.52	0.58	0.42	0.51
Asian	0.66	0.72	0.56	0.65
Hispanic	0.60	0.66	0.50	0.58

Appendix 3: Using the R Code

The file `functions.r` includes several functions that allow us to calculate the measures proposed in this paper. To load these functions, it is possible to use `source("PATH/functions.r")`, where `PATH` has to be replaced with the path where the file `functions.r` is stored, e.g. `C:/Code`.

The main function, which is called `pod`, calculates the indices presented in this paper. In addition, several other functions are available to assess the results, including a function for the estimation of confidence intervals based on bootstrapping. Brief comments on the usage of the functions are given in the file `functions.r`. Another file, `example.r`, includes short examples on the usage of the functions based on NHIS data, which can be downloaded at no cost from <https://www.cdc.gov/nchs/nhis/> (Tables 8, 9).

The `pod` function is used as

```
pod(data, groups, subgroups = NULL, measures = NULL, weights = NULL)
```

where the arguments can be used as follows:

data	This argument specifies which data set should be used for analysis. The data have to be of class <code>data.frame</code> or <code>matrix</code> . If, for instance, the user wants to analyze a data set that is stored under the name <code>healthdata</code> , then this argument would be specified as <code>data=healthdata</code> .
groups	This argument specifies a variable name included in the data set given by <code>data</code> . The variable is used to define the groups that are being compared. More specifically, each unique value of the variable will be assumed to represent one group. Note that the variable name has to be given in parentheses. For example, if the data set includes a variable called <code>gender</code> , the argument would be specified as <code>groups="gender"</code> .
subgroups	This argument is optional and can be used to specify the name of a variable that gives subgroups for decomposition. It works like the argument <code>groups</code> . If it is specified, a decomposition of the PODI by subgroups will be calculated.
measures	This argument can be used to specify the name(s) of variables giving information on conditions. These variables have to be binary. If this argument is not specified, it is assumed that all variables included in the data give information on conditions, except those variables used by other arguments (<code>groups</code> , <code>subgroups</code> , <code>eights</code>). For example, if we want to specify this argument using three variables called <code>condition1</code> , <code>condition2</code> , and <code>condition3</code> , the argument would be used as <code>measures=c("condition1", "condition2", "condition3")</code> .
weights	This argument is optional and can be used to specify the name of a variable giving survey weights.

Taking the above examples together a call of the `poi` function might look like

```
poi(data = healthdata, groups = egendere)
```

This would use the data set `healthdata` and would run the analysis by `gender`. As the `measures` are not specified, all of the variables in `healthdata` except `gender` are assumed to contain information on conditions. Several other examples that illustrate how to use this functions and how to view its results are given in the file `example.r`.

References

- Abraido-Lanza, A. F., Dohrenwend, B. P., Ng-Mak, D. S., & Turner, J. B. (1999). The latino mortality paradox: a test of the "salmon bias" and healthy migrant hypotheses. *American Journal of Public Health*, 89(10), 1543–1548.
- Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, 95, 476–487.
- Anderson, G. F., Jeremy, H., Hussey, P. S., & Jee-Hughes, M. (2000). Health spending and outcomes: Trends in oecd countries, 1960–1998. *Health Affairs*, 19, 150–157.
- Arndt, C., Distante, R., Hussain, M. A., Østerdal, L. P., Huong, P. L., & Ibraimo, M. (2012). Ordinal welfare comparisons with multiple discrete indicators: A first order dominance approach and application to child poverty. *World Development*, 40, 2290–2301.
- Bago d'Uva, T., van Doorslaer, E., & Lindeboom, M. (2008). Does reporting heterogeneity bias the measurement of health disparities? *Health Economics*, 17, 351–375.
- Banks, J., Marmot, M., Oldfield, Z., & Smith, J. P. (2006). Disease and disadvantage in the United States and in England. *Journal of the American Medical Association*, 295, 2037–2045.

- Barzilai, J., & Golany, B. (1990). Deriving weights from pairwise comparison matrices: The additive case. *Operations Research Letters*, 9, 407–410.
- Braveman, P. A., Cubbin, C., Egerter, S., Williams, D. R., & Pamuk, E. (2010). Socioeconomic disparities in health in the United States: What the patterns tell us. *American Journal of Public Health*, 100, S186–S196.
- Brilleman, S. L., Gravelle, H., Hollinghurst, S., Purdy, S., Salisbury, C., & Windmeijer, F. (2014). Keep it simple? Predicting primary health care costs with clinical morbidity measures. *Journal of Health Economics*, 35, 109–122.
- Bruggemann, R., & Carlsen, L. (2011). An improved estimation of averaged ranks of partial orders. *MATCH Communications in Mathematical and in Computer Chemistry*, 65, 383–414.
- Burgard, S. A., & Chen, P. V. (2014). Challenges of health measurement in studies of health disparities. *Social Science & Medicine*, 106, 143–150.
- Charlson, M. E., Pompei, P., Ales, K. L., & MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases*, 40, 373–383.
- Crawford, G., & Williams, C. (1985). A note on the analysis of subjective judgment matrices. *Journal of Mathematical Psychology*, 29, 387–405.
- de Groot, V., Beckerman, H., Lankhorst, G. J., & Bouter, L. M. (2003). How to measure comorbidity: A critical review of available methods. *Journal of Clinical Epidemiology*, 56, 221–229.
- Devlin, N. J., & Brooks, R. (2017). EQ-5D and the EuroQol group: Past, present and future. *Applied Health Economics and Health Policy*, 15, 127–137.
- Devlin, N. J., Shah, K. K., Feng, Y., Mulhern, B., & van Hout, B. (2018). Valuing health-related quality of life: An EQ-5 D-5 l value set for England. *Health Economics*, 27, 7–22.
- Dowd, J. B., & Zajacova, A. (2010). Does self-rated health mean the same thing across socioeconomic groups? Evidence from biomarker data. *Annals of Epidemiology*, 20(10), 743–749.
- Etches, V., Frank, J., Di Ruggiero, E., & Manuel, D. (2006). Measuring population health: A review of indicators. *Annual Review of Public Health*, 27, 29–55.
- Etilé, F., & Milcent, C. (2006). Income-related reporting heterogeneity in self-assessed health: Evidence from France. *Health Economics*, 15(9), 965–981.
- Fattore, M. (2016). Partially ordered sets and the measurement of multidimensional ordinal deprivation. *Social Indicators Research*, 128(2), 835–858.
- Fattore, M., & Arcagni, A. (2018). A reduced poset approach to the measurement of multidimensional ordinal deprivation. *Social Indicators Research*, 136(3), 1053–1070.
- Fattore, M., & Bruggemann, R. (Eds.). (2017). *Partial order concepts in applied sciences*. Berlin: Springer.
- Fattore, M., & Maggino, F. (2015). A new method for measuring and analyzing suffering: Comparing suffering patterns in Italian society. In R. E. Anderson (Ed.), *World suffering and quality of life* (pp. 385–399). Berlin: Springer.
- Fattore, M., Maggino, F., & Greselin, F. (2011). Socio-economic evaluation with ordinal variables: Integrating counting and poset approaches. *Statistica & Applicazioni Special Issue*, 1, 31–42.
- Gavalec, M., Ramík, J., & Zimmermann, K. (2015). *Decision making and optimization. Special matrices and their applications in economics and management. Lecture notes in economics and mathematical systems*. New York: Springer.
- Hallstrom, A. P., Cobb, L. A., & Yu, B. H. (1996). Influence of comorbidity on the outcome of patients treated for out-of-hospital ventricular fibrillation. *Circulation*, 93, 2019–2022.
- Herrero, C., & Villar, A. (2013). On the comparison of group performance with categorical data. *PLoS ONE*, 8(e84), 784.
- Hummer, R. A., Benjamins, M. R., & Rogers, R. G. (2004). Racial and ethnic disparities in health and mortality among the US elderly population. In N. B. Anderson, R. A. Bulato, & B. Cohen (Eds.), *Critical perspectives on racial and ethnic differences in health in late life* (pp. 53–94). Washington, DC: National Academy Press.
- Humphries, K. H., & Van Doorslaer, E. (2000). Income-related health inequality in Canada. *Social Science & Medicine*, 50(5), 663–671.
- Hurwitz, E. L., & Morgenstern, H. (1997). The effects of comorbidity and other factors on medical versus chiropractic care for back problems. *Spine*, 22, 2254–2263.
- Hussain, M. A., Jørgensen, M. M., & Østerdal, L. P. (2016). Refining population health comparisons: A multidimensional first order dominance approach. *Social Indicators Research*, 129, 739–759.
- Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior*, 38, 21–37.
- Idler, E. L., & Benyamini, Y. (1999). Community studies reporting association between self-rated health and mortality. *Research on Aging*, 21, 392–401.

- Johnston, D. W., & Propper, M. A. C. S. (2009). Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient. *Journal of Health Economics*, 28(3), 540–552.
- Jürges, H. (2008). Self-assessed health, reference levels and mortality. *Applied Economics*, 40(5), 569–582.
- König, H. H., Bernert, S., Angermeyer, M. C., Matschinger, H., Martinez, M., Villagut, G., et al. (2009). Comparison of population health status in six European countries: Results of a representative survey using the EQ-5D questionnaire. *Medical Care*, 47, 255–261.
- Lieberman, S. (1976). Rank-sum comparison between groups. *Sociological Methodology*, 7, 276–291.
- Marengoni, A., Winblad, B., Karp, A., & Fratiglioni, L. (2008). Prevalence of chronic diseases and multimorbidity among the elderly population in Sweden. *American Journal of Public Health*, 98, 1198–1200.
- Marmot, M. (2005). Social determinants of health inequalities. *The Lancet*, 365, 1099–1104.
- Morton, J., Pachter, L., Shiu, A., Sturmfels, B., & Wienand, O. (2009). Convex rank tests and semigraphoids. *SIAM Journal on Discrete Mathematics*, 23, 1117–1134.
- Murray, C. J. L. (2002). *Summary measures of population health: Concepts, ethics, measurement and applications*. Geneva: World Health Organization.
- National Research Council of the National Academics. (2004). *Critical perspectives on racial and ethnic differences in health in late life*. Washington: National Academies Press.
- Permanyer, I., & Hussain, M. A. (2018). First order dominance techniques and multidimensional poverty indices: An empirical comparison of different approaches. *Social Indicators Research*, 137, 867–893.
- Rivera, J. A., Barquera, S., Campirano, F., Campos, I., Safdie, M., & Tovar, V. (2002). Epidemiological and nutritional transition in Mexico: Rapid increase of non-communicable chronic diseases and obesity. *Public Health Nutrition*, 5, 113–122.
- Schiøtz, M. L., Stockmarr, A., Høst, D., Glümer, C., & Frølich, A. (2017). Social disparities in the prevalence of multimorbidity—A register-based population study. *BMC Public Health*, 17, 422.
- Stewart, A. L., Ware, J. E., & Ware, J. E, Jr. (Eds.). (1992). *Measuring functioning and well-being: The medical outcomes study approach*. Durham: Duke University Press.
- Van Doorslaer, E., & Gerdtham, U. G. (2003). Does inequality in self-assessed health predict inequality in survival by income? Evidence from Swedish data. *Social Science & Medicine*, 57(9), 1621–1629.
- Van Doorslaer, E., Koolman, X., & Jones, A. M. (2002). Explaining income-related inequalities in doctor utilisation in Europe: A decomposition approach, ECuity II Project
- Ward, B. W. (2013). Prevalence of multiple chronic conditions among US adults: Estimates from the national health interview survey, 2010. *Preventing Chronic Disease*, 10, E65.
- Weimann, A., Dai, D., & Oni, T. (2016). A cross-sectional and spatial analysis of the prevalence of multimorbidity and its association with socioeconomic disadvantage in South Africa: A comparison between 2008 and 2012. *Social Science & Medicine*, 163, 144–156.